

135042
Shears, Beverly

From: Devi, Sarvamangala
Sent: Tuesday, October 12, 2004 11:10 AM
To: Shears, Beverly
Subject: 10/625,972

Good morning Beverly:

In application 10/625,972, would please performe a sequence search for SEQ ID NO: 4 in commercial and interference databases. Please also run SEQ ID NO: 4 amino acid databases.

Please include an inventors' name search: Phillip I. Tarr; Sima S. Bilge; Thomas E. Besser; James C. Vary

Thanks.

S. DEVI, Ph.D.
AU 1645
Rems - 3C18



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 135042

To: Sarvamangala Devi
Location: REM 3C18
Art Unit: 1645
Friday, October 15, 2004
Case Serial Number: 10/625972

From: Beverly Shears
Location: Remsen Bldg.
RM 1A54
Phone: 571-272-2528
beverly.shears@uspto.gov

Search Notes

Date completed: _____
Searcher: Beverly e 2528
Terminal time: _____
Elapsed time: _____
CPU time: _____
Total time: _____
Number of Searches: _____
Number of Databases: _____

Search Site
____ STIC
____ CM-1
____ Pre-S
Type of Search
____ N.A. Sequence
____ A.A. Sequence
____ Structure
____ Bibliographic

Vendors
____ IG
____ ☒ STN
____ ☒ Dialog
____ APS
____ Geninfo
____ SDC
____ DARC/Questel
____ ☒ Other CGN

This Page Blank (uspto)

GenCore version 5.1.6 -
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: October 13, 2004, 11:44:12 ; Search time 26.5 Seconds
(without alignments)
8147.163 Million cell updates/sec

Title: US-10-625-972-4
Perfect score: 3732
Sequence: 1 atcggaataaccactctggc.....cgctgaactatcagttctga 2091

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Zgapop 6.0 , Zgapext 7.0
Delop 6.0 , Delext 7.0

389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 778828

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10625972/runat_12102004_130106_26853/app.query.fasta_1.2247
-DB=Issued Patents AA -QMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MDSB=LOCAL -OUTFWT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10625972@cgn_1_25 @runat_12102004_130106_26853 -NCPU=6 -ICPU=3
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA:
1: /cgn2_6/ptodata/2/iaa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PTCUS.COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3638	97.5	696	1	US-08-765-081-5
2	3638	97.5	696	3	US-09-098-082-5
3	3638	97.5	696	5	PCT-US95-06994-5
4	3594.5	96.3	703	5	PCT-US95-06994-8
5	3557	95.3	718	5	PCT-US95-06994-5
6	1089.5	29.2	718	5	PCT-US95-06994-7
7	1052.5	28.2	652	1	US-08-765-081-6
8	1052.5	28.2	652	3	US-09-098-082-6
9	1007.5	27.0	677	4	US-09-543-681A-7936
10	956.5	25.6	684	4	US-09-543-681A-4908
11	880.5	23.6	699	4	US-09-489-039A-8133
12	805.5	21.6	703	5	PCT-US95-06994-9

13	800.5	21.4	663	1	US-08-765-081-7
14	800.5	21.4	663	3	US-09-098-082-7
15	682	18.3	778	4	US-09-489-039A-8455
16	642.5	17.2	759	4	US-09-489-039A-7874
17	641	17.2	795	4	US-09-489-039A-13394
18	617.5	16.5	761	4	US-09-328-352-5942
19	607.5	16.3	478	4	US-09-252-991A-31866
20	604.5	16.2	725	4	US-09-668-113A-8
21	602	16.1	744	4	US-09-252-991A-29565
22	537	14.4	651	4	US-09-543-681A-8284
23	400	10.7	617	4	US-09-252-991A-30921
24	396.5	10.6	1086	4	US-09-252-991A-25051
25	383.5	10.3	643	4	US-09-328-352-5146
26	371.5	10.0	498	4	US-09-489-039A-12187
27	360.5	9.7	478	4	US-09-252-991A-31603
28	359.5	9.6	789	4	US-09-252-991A-27011
29	349	9.4	704	4	US-09-489-039A-9841
30	340	9.1	704	4	US-09-328-352-5537
31	339.5	9.1	695	4	US-09-489-039A-14338
32	316.5	8.5	803	4	US-09-252-991A-28469
33	315	8.4	685	4	US-09-543-681A-4482
34	315	8.4	743	4	US-09-328-352-5073
35	312.5	8.4	790	3	US-08-817-707-6
36	311.5	8.3	718	4	US-09-252-991A-29432
37	311	8.3	844	4	US-09-252-991A-25037
38	307.5	8.2	727	4	US-09-489-039A-14168
39	304	8.1	791	3	US-08-537-361E-4
40	303	8.1	903	4	US-09-252-991A-24977
41	301	8.1	790	3	US-08-817-707-8
42	301	8.1	791	3	US-08-817-707-4
43	298.5	8.0	792	3	US-08-537-361E-2
44	294.5	7.9	766	4	US-09-328-352-8230
45	293.5	7.9	792	1	US-08-326-670A-2

ALIGNMENTS

RESULT 1

US-08-765-081-5
Sequence 5, Application US/08765081
Patent No. 5798260
GENERAL INFORMATION:
APPLICANT: Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
TITLE OF INVENTION: Escherichia Coli 0157:H7 Epithelial Adhesin
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
STREET: 2800 Pacific First Center, 1420 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98101-2347
COMPUTER READABLE FORM:
SEQUENCE TYPE: Diskette-3.5 inch, 1.44Mb storage
COMPUTER: IBM PC compatible/Pentium
OPERATING SYSTEM: MS-Windows 3.1
SOFTWARE: Word for Windows-6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/765,081
FILING DATE: March 26, 1997
PRIOR APPLICATION DATA: PCT/US95/06994
APPLICATION NUMBER: PCT/US95/06994
FILING DATE: June 7, 1995
APPLICATION NUMBER: US 08/265,714
FILING DATE: June 24, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Shelton, Dennis K.
REGISTRATION NUMBER: 26,997
REFERENCE/DOCKET NUMBER: CHOR-1-10286
TELECOMMUNICATION INFORMATION:
TELEPHONE: 1-206-682-8100; 1-206-224-0718 (direct)
TELEFAX: 1-206-224-0779
INFORMATION FOR SEQ ID NO: 5:

Sequence 7, Appli
Sequence 7, Appli
Sequence 8455, Ap
Sequence 7874, Ap
Sequence 13394, A
Sequence 5942, Ap
Sequence 31866, A
Sequence 8, Appli
Sequence 29565, A
Sequence 8284, Ap
Sequence 30921, A
Sequence 25051, A
Sequence 5146, Ap
Sequence 12187, A
Sequence 31603, A
Sequence 27011, A
Sequence 9841, Ap
Sequence 5537, Ap
Sequence 14338, A
Sequence 28469, A
Sequence 4482, Ap
Sequence 5073, Ap
Sequence 6, Appli
Sequence 29432, A
Sequence 25037, A
Sequence 14168, A
Sequence 24977, A
Sequence 4, Appli
Sequence 8, Appli
Sequence 2, Appli
Sequence 8230, Ap
Sequence 2, Appli

SEQUENCE CHARACTERISTICS:
 ; LENGTH: 696 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-765-081-5

Alignment Scores:

Pred. No.: 0 Length: 696
 Score: 3638.00 Matches: 696
 Percent Similarity: 100.00% Conservatives: 0
 Best local Similarity: 100.00% Mismatches: 0
 Query Match: 97.48% Indels: 0
 DB: 1 Gaps: 0

US-10-625-972-4 (1-2091) x US-08-765-081-5 (1-696)

```

QY 1 ATCGGAATAACCACTCTGGCTTCGGTAGTCATTCCTGTCTCGGATTTTCAGCCAGCAGC 60
DB 1 MetArgIleThrThrLeuAlaSerValValIleProCysLeuGlyPheSerAlaSerSer 20
QY 61 ATAGCTGCTGCAGAGGATGTGATGTTGTCGCGCATCCGGCTATGAGAAAAAGCTGACT 120
DB 21 IleAlaAlaAlaGluAspValMetIleValSerAlaSerGlyTyrGluLysLysLeuThr 40
QY 121 AACCGAGCCGACAGTGTTCCTGTGATTAGCCAGGAGGAATTCAGTCCAGCCAGTACCAC 180
DB 41 AsnAlaAlaAlaSerValSerValIleSerGlnGluLeuGlnSerSerGlnTyrHis 60
QY 181 GATCTCGCGGAGCTCTGAGATCAGTAGAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 240
DB 61 AspLeuAlaGluAlaLeuArgSerValGluGlyValAspValGluSerGlyThrGlyLys 80
QY 241 ACCGAGGCTCGAATCAGATCCAGGAATGCCAGCCAGGTACACCTGATCTACTGATT 300
DB 81 ThrGlyGlyLeuGluLysSerIleArgGlyMetProAlaSerTyrThrLeuIleLeuIle 100
QY 301 GATGTTGTCGAGGCGGAGCAGTACGCTGACTCCCAAGCTTTCTGCCATGAT 360
DB 101 AspGlyValArgGlnGlySerSerAspValThrProAsnGlyPheSerAlaMetAsn 120
QY 361 ACCGGTTTCATCCCTCTGCGCCCATTTGAGCGTATTGAGGTTATCAGGGGGCCGATG 420
DB 121 ThrGlyPheMetProProLeuAlaAlaIleGluArgIleGluValIleArgGlyProMet 140
QY 421 TCCACTGTATGGCTCTGATCGGATGGCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 480
DB 141 SerThrLeuTyrGlySerAspAlaMetGlyGlyValValAsnIleIleThrArgLysAsn 160
QY 481 GCAGACAAATGGCTCTCTCCGTCAATCAGGGCTGAACTTCGAGGAAAGCAACAATGG 540
DB 161 AlaAspLysTrpLeuSerSerValAsnAlaGlyLeuAsnLeuGlnGluSerAsnLysTrp 180
QY 541 GGTACAGCAGCAGTAAATTTCTGGAGCAGTGTGCTCCCTTGTGGATGATCTGTGAGC 600
DB 181 GlyAsnSerSerGlnPheAsnPheTrpSerSerGlyProLeuValAspSerValSer 200
QY 601 CTGCAAGTACGGGTAGCACACAACAGCGTCAGGGTTCATCGGTTCACATCACTCAGCGAT 660
DB 201 LeuGlnValArgGlySerThrGlnArgGlnGlySerSerValThrSerLeuSerAsp 220
QY 661 ACAGCAGCAGCGTATTCCTTATCCACGAGTACAGAAATTAATCTTGTGGTGCAGT 720
DB 221 ThrAlaGlyThrArgIleProTyrProThrGluSerGlnAsnTyrAsnLeuGlyAlaArg 240
QY 721 CTTGACTGGAAGCTCGGACAGGATGTCTGTTGTATGATATGATATACCCCGCAG 780
DB 241 LeuAspTrpLysAlaSerGluGlnAspValLeuTrpPheAspMetAspThrThrArgGln 260
QY 781 COTTATGATAACCGGATGGCAACTGGGGATCTGACGGGGGATATGACCGGACCCCTG 840
DB 261 ArgTyrAspAsnArgAspGlyGlnLeuGlySerLeuThrGlyGlyTyrAspArgThrLeu 280

```

```

QY 841 CGCTATCAGCGAAACAAAATTTTCAGCTGGCTATGATCATCTTTACCTTCGGAAACATGG 900
DB 281 ArgTyrGluArgAsnLysIleSerAlaGlyTyrAspHisThrPheThrPheGlyThrTrp 300
QY 901 AAATCGTATCTGAACCTGGAAACGAGACAGAAAAATAAGGTGCTGAGCTTTGACCATGTA 960
DB 301 LysSerTyrLeuAsnTrpAsnGluThrGluAsnLysGlyArgGluLeuValArgSerVal 320
QY 961 CTGAAGCGCACAAATGGGGCTTCGGTFCAGCCCGGGAGCTTAAGAAATCGAACCTT 1020
DB 321 LeuLysArgAspLysTrpGlyLeuAlaGlyInProArgLuuLeuLysGluSerAsnLeu 340
QY 1021 ATCTGAAATTCATTACTGCTTACCCCTCTCGGAGAAATCTCATCTGCTTACGTTGGGGG 1080
DB 341 IleLeuAsnSerLeuLeuLeuThrProLeuGlyGluSerHisLeuValThrValGlyGly 360
QY 1081 GAGTTTCAGAGCTCGCTCATGAAGAGCGAGTTGCTTCCAGACAGAGTGAATTTTC 1140
DB 361 GluPheGlnSerSerSerMetLysAspGlyValValLeuAlaSerThrGlyGluThrPhe 380
QY 1141 CGCAGAAAAAGCTGGTGGTATTTGCTGAGGATGAGTGGCATCTCACGGATGACACTTGG 1200
DB 381 ArgGlnLysSerTrpSerValPheAlaGluAspGluTrpHisLeuThrAspAlaLeuAla 400
QY 1201 CTGACTCGCGGACGCGCTATGAACATCATGAGCAATTCGGGGGACACTTCAGTCCGGCT 1260
DB 401 LeuThrAlaGlySerArgTyrGluHisHisGluGlnPheGlyGlyHisPheSerProArg 420
QY 1261 GCATATCTGCTGGGATGTGGCAGATGCTGACGCTCAAAAGGCGGTGTGACCCAGGGA 1320
DB 421 AlaTyrLeuValTrpAspValAlaAspAlaTrpThrLeuLysGlyGlyValThrGly 440
QY 1321 TATAAGCCACCCAGATGGGCGAGCTACATAAAGGATTAGTGGTGTGTCCGGCAGGGA 1380
DB 441 TyrLysAlaProArgMetGlyGlnLeuHisLysGlyIleSerGlyValSerGlyGlnGly 460
QY 1381 AAACAAATCTACTCTGTAACCCCGACCTGAAGCGGAGAGAGAGCGTCAAGTTATGAGCT 1440
DB 461 LysThrAsnLeuLeuGlyAsnProAspLeuLysProGluGluSerValSerTyrGluAla 480
QY 1441 GGGGTCTATTACGATAACCCCGCTCTGAAATGTCGAATGTCCACAGTTTATGACTGAC 1500
DB 481 GlyValTyrTyrAspAsnProAlaGlyLeuAsnAlaAsnValThrGlyPheMetThrAsp 500
QY 1501 TTCTCCAAACAGANTTCTCTTTATCCATAATATGATACACCAATAGCTATGTAACAGC 1560
DB 501 PheSerAsnLysIleValSerTyrSerIleAsnAspAsnThrAsnSerTyrValAsnSer 520
QY 1561 GAAAAGGCCCGGTTCAGCGGTGTGAAATTTGCCGCGCACATTCGCCCTGTGTGTCAGAGAT 1620
DB 521 GlyLysAlaArgLeuHisGlyValGluPheAlaGlyThrLeuProLeuTrpSerGluAsp 540
QY 1621 GTCACTCTGCTGACTGAATTTACACCTGGACCCGAGTGAACAACAGTATGCTGATACAAA 1680
DB 541 ValThrLeuSerLeuAsnTyrThrTrpThrArgSerGluGlnArgAspGlyAspAsnLys 560
QY 1681 GGTGGCCCTCAGTTATACCCCTGAAACACATGTTGAATGCGAACTGAACCTGGCAGATC 1740
DB 561 GlyAlaProLeuSerTyrThrProGluHisMetValAsnAlaLysLeuAsnTrpGlnIle 580
QY 1741 ACCGAAGAGGTGGCATCATGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1800
DB 581 ThrGluGluValAlaSerTrpLeuGlyAlaArgTyrArgLysThrProArgPheThr 600
QY 1801 CAGAATTATTCGTCAGCTGAGCGCTGTACAGAGAAGTGTATGATGAGAAAGGACAATAC 1860
DB 601 GlnAsnTyrSerSerLeuSerAlaValGlnLysLysValTyrAspGluLysGlyGlyLys 620
QY 1861 CTCAAAGCTGGACGGTGGTGGATCGAGTCTGTCTGTGAAAGATGACGGATGCCCTGACG 1920
DB 621 LeuLysAlaTrpThrValValAspAlaGlyLeuSerTrpLysMetThrAspAlaLeuThr 640
QY 1921 CTGAATGCTCGCGTGAATAACCTGCTCAACAGGATTACAGTGCAGCTGACGCTGTACAGT 1980

```


Db 641 LeuAsnAlaAlaValAsnAsnLeuLeuAsnLysAspTyrSerAspValSerLeuTyrSer 560
Qy 1981 GCCGCTAAGAGTACGCTGTATGCGGTGATTACTTCCAGACGGGATCATCAACACAGCA 2040
Db 561 AlAGlyLysSerThrLeuTyrAlaGlyAspTyrPheGlnThrGlySerSerThrThrGly 680
Qy 2041 TATGTGATCTAGCGAATAATTACTGATGTCCTGACATCATCAGTTC 2088
Db 681 TyrValIleProGluArgAsnTyrTrpMetSerLeuAsnTyrGlnPhe 696
RESULT 2
US-09-098-082-5
; Sequence 5, Application US/09098082
; Patent No. 6040421
; GENERAL INFORMATION:
; APPLICANT: Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
; TITLE OF INVENTION: Escherichia Coli 0157:H7 Epithelial Adhesin
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
; STREET: 2800 Pacific First Center, 1420 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette-3.5 inch, 1.44Mb storage
; COMPUTER: IBM PC compatible/Pentium II
; OPERATING SYSTEM: MS-Windows 95
; SOFTWARE: Word for Windows-6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/098,082
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/765,081
; FILING DATE: March 26, 1997
; APPLICATION NUMBER: PCT/US95/06994
; FILING DATE: June 7, 1995
; APPLICATION NUMBER: US 08/265,714
; FILING DATE: June 24, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Shelness, Diana K.
; REGISTRATION NUMBER: 35,356
; REFERENCE/DOCKET NUMBER: CHOR-1-12402
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 1-206-682-8100; 1-206-224-0735 (direct)
; TELEFAX: 1-206-224-0779
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 696 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-098-082-5
Alignment Scores:
Pred. No.: 0 Length: 696
Score: 3638.00 Matches: 696
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 97.48% Indels: 0
DB: 3 Gaps: 0
US-10-625-972-4 (1-2091) x US-09-098-082-5 (1-696)
Qy 1 ATCGGAATACCACTCTGCTTCGCTAGTTCCTGCTCGGATTTTCAGCCAGCAGC 60
Db 1 MetArgIleThrLeuAlaSerValValIleProCysLeuGlyPheSerAlaSerSer 20
Qy 61 ATAGCTGCTGACAGAGTGTGATGTTCTCGGATCCGGTATGAGAAAGCTGACT 120
Db 21 IleAlaAlaAlaGluAspValMetIleValSerAlaSerGlyTyrGluLysLeuThr 40

Qy 121 AACGACGCCGCGAGTGTTCCTGTGATTAGCCAGGAGGAATTGCGATCCAGCCAGTACCAC 180
Db 41 AsnAlaAlaAlaSerValSerValIleSerGlnGluLeuGlnSerSerGlnTyrHis 60
Qy 181 GATCTGGCGGAGGCTCTCAGATCAGTAGAGGTGTGGATGTGAAAGTGGTACGGGTAAA 240
Db 61 AspLeuAlaGluAlaLeuArgSerValGluGlyValAspValGluSerGlyThrGlyLys 80
Qy 241 ACCGAGGCGTGGAAATCAGATCCGAGGAATGCCAGCCAGTTACACGCTGATACATGATT 300
Db 81 ThrGlyGlyLeuGluIleSerIleArgGlyMetProAlaSerTyrThrLeuIleLeuIle 100
Qy 301 GATGTGTTCGTCCAGGCGGAGACAGTGCAGTACTCCCAACGGTTCCTCCCATGAAAT 360
Db 101 AspGlyValArgGlnGlyGlySerSerAspValThrProAsnGlyPheSerAlaMetAsn 120
Qy 361 ACCGGTTCATCCCTCTGCTGCGCCCATTTGAGCGTATTGAGGTATCAGGGGGCCGATG 420
Db 121 ThrGlyPheMetProProLeuAlaIleGluArgIleGluValIleArgGlyProMet 140
Qy 421 TCCACACTGTATGGCTCTGATCGGATGGCGGTGTGTGAATATCATCACCAGAAAGAT 480
Db 141 SerThrLeuTyrGlySerAspAlaMetGlyGlyValValAsnIleIleThrArgLysAsn 160
Qy 481 GCAGACAAATGGCTCTCTTCCTCAATGAGGCTGAATCTCCAGAAAGCAACAAATGG 540
Db 161 AlaAspLysTrpLeuSerSerValAsnAlaGlyLeuAsnLeuGlnGluSerAsnLysTrp 180
Qy 541 GGTACAGCAGCAGTCTTAATTTCTGGAGCAGTGTCCCTTTGTGATGATTCTGTGACG 600
Db 181 GlyAsnSerSerGlnPheAsnPheTrpSerSerGlyProLeuValAspSerValSer 200
Qy 601 CTCAGGTACCGGTAGCACACAGCCTCAGGTTTCATCGGTCAATCATCAGTACATCAGCAG 560
Db 201 LeuGlnValArgGlySerThrGlnArgGlnGlySerSerValThrSerLeuSerAsp 220
Qy 661 ACAGCAGCAGCAGTCTTCTTATCCACGAGTACACAGAAATATATCTTCGTGACGT 720
Db 221 ThrAlaGlyThrArgIleProTyrProThrGluSerGlnAsnTyrAsnLeuGlyAlaArg 240
Qy 721 CTGACTGGAAGCGCTCGGAGCAGGATGTGCTCTGTTTGTATGATACCAACCCGCGAG 780
Db 241 LeuAspTrpLysAlaSerGluGlnAspValLeuTrpPheAspMetAspThrThrArgGln 260
Qy 781 CGTTATGATAACCGGATGGCACTGGGAGTCTGACCGGGGATATCAGCGGACCCCTG 840
Db 261 ArgTyrAspAsnArgAspGlyGlnLeuGlySerLeuThrGlyGlyTyrAspArgThrLeu 280
Qy 841 CGCTATGAGCGAACAATAATTTTCAGCTGGCTATGATCATCTTACCTTCGGAACATGG 900
Db 281 ArgTyrGluArgAsnLysIleSerAlaGlyTyrAspHisThrPheThrPheGlyThrTrp 300
Qy 901 AAATCGTATCTGAATCGAAGACGAGACAGAAATAAAGGTCTGTGAGCTGTACCGAGTGA 960
Db 301 LysSerTyrLeuAsnTrpAsnGluThrGluAsnLysGlyArgGluLeuValArgSerVal 320
Qy 961 CTGAACGCGACAAATGGGGCTTCGCGTCCAGCCGCGAGCTTAAGGAATCGAACCTT 1020
Db 321 LeuLysArgAspLysTrpGlyLeuAlaGlyGlnProArgGluLeuLysGluSerAsnLeu 340
Qy 1021 ATCTGAAATTCATTACTGCTTACCCCTCTGGGAGAATCTCATCTGTTACGTTGGGGGGC 1080
Db 341 IleLeuAsnSerLeuLeuLeuThrProLeuGlyGluSerHisLeuValThrValGlyGly 360
Qy 1081 GAGTTTCAGAGCTCGTCCATGAAGACGAGGATTCCTCTCCAGCAGCAGGTGAACCTTC 1140
Db 361 GluPheGlnSerSerMetLysAspGlyValValLeuAlaSerThrGlyGluThrPhe 380
Qy 1141 CGGCAGAAAAGCTGCTCGGTATTTGCTGAGGATCAGTGGCATCTCAGCGATGCACTTGG 1200
Db 381 ArgGlnLysSerTrpSerValPheAlaGluAspGluTrpHisLeuThrAspAlaLeuAla 400

1201 CTGACTCGGGCAGCCGCTATGACATCATGACCAATTCGGGGACACCTTCAGTCGGCT 1260
 401 LeuThrAlaGlySerArgTyrGluHisHisGluGlnPheGlyHisPheSerProArg 420
 1261 GCATATCTGCTCTGGGATGGCAGATGCTGACCGCTGAAGCGGCTGTGACCGGA 1320
 421 AlaTyrLeuValTrpAspValAlaAspAlaTrpThrLeuLysGlyGlyValThrThrGly 440
 1321 TATAAGCACCAGATGGGCGAGCTACATAAAGGATAGTGGTGTCGGGCGAGGA 1380
 441 TyrLysAlaProArgMetGlyGlnLeuHisLysGlyLysGlyValSerGlyGlnGly 460
 1381 AAAACAAATCTACTTGGTAAACCCGACCTGAAGCCGAGAGAGCGCTAGTTATGAGCT 1440
 461 LysThrAsnLeuLeuGlyAsnProAspLeuLysProGluLysValSerTyrGluAla 480
 1441 GGGGTGTATTAAGTAACCCCGCTGTAATGCCAATGTACAGGTTTATGACTGAC 1500
 481 GlyValTyrTyrAspAsnProAlaGlyLeuAsnAlaAsnValThrGlyPheMetThrAsp 500
 1501 TTCTCCACAGAGATGCTCTTATTCATATAATGATACCAATAGCTATGTAACAGC 1560
 501 PheSerAsnLysIleValSerTyrSerIleAsnAspAsnThrAsnSerTyrValAsnSer 520
 1561 GGAAGGCCCGGTGACCGGTGGAATTTGCCGCGACATTCGCCGCTGTGTGTCAGAGAT 1620
 521 GlyLysAlaArgLeuHisGlyValGluPheAlaGlyThrLeuProLeuTrpSerGluAsp 540
 1621 CTCACGCTGCTACTGATTAACCTGGACCGCCGAAGTGAACACAGTGATGGTGATACAAA 1680
 541 ValThrLeuSerLeuAsnTyrThrTrpThrArgSerGluGlnArgAspGlyAspAsnLys 560
 1681 GGTGCCCGCTGAGTTATACCCCTGAACACATGTTGAATGCGAACTGAACCTGGCAGATC 1740
 561 GlyAlaProLeuSerTyrThrProGluHisMetValAsnAlaLysLeuAsnTrpGlnIle 580
 1741 ACCGAAGAGTGGCATCATGCTGGTGGTCCCTGTTATCGGGGAAACACACGTTTCACC 1800
 581 ThrGluGluValAlaSerTrpLeuGlyAlaArgTyrArgGlyLysThrProArgPheThr 600
 1801 CAGAATTATTCGCTACTGAGCGCTGTACAGAGAAGTGTATGATGAGAAGGAGCAATAC 1860
 601 GlnAsnTyrSerSerLeuSerAlaValGlnLysLysValTyrAspGluLysGlyLysThr 620
 1861 CTGAAGCTGACCGGTGGATGATGCTGCTGGAAGATGACGATGCCCTGACG 1920
 621 LeuLysAlaTrpThrValValAlaGlyLeuSerTrpLysMetThrAspAlaLeuThr 640
 1921 CTGAATGCTGCGGTGAATACCTGCTCAACAGGATTAACGTCGCTGAGCGCTGTACAGT 1980
 641 LeuAsnAlaAlaValAsnAsnLeuLeuAsnLysAspTyrSerAspValSerLeuTyrSer 660
 1981 GCCGTGAAGATGACGCTGTATGCCGCTGATTAATCTCCAGCGGATCATCAACACAGGA 2040
 661 AlaGlyLysSerThrLeuTyrAlaGlyAspTyrPheGlnThrGlySerSerThrThrGly 680
 2041 TATGTGATACCTGACGGAATTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2088
 681 TyrValIleProGluArgAsnTyrTrpMetSerLeuAsnTyrGlnPhe 696

RESULT 3

PCT-US95-06994-5

; Sequence 5: Application PC/TUS9506994

; GENERAL INFORMATION:

; APPLICANT: Children's Hospital & Medical Center

; APPLICANT: University of Washington

; APPLICANT: Washington State University Research Foundation

; APPLICANT: TARR, PHILLIP I

; APPLICANT: BILGE, SIMA S

; APPLICANT: BESSER, THOMAS E

; APPLICANT: VARY JR, JAMES C

; TITLE OF INVENTION: ESCHERICHIA COLI O157:H7 EPITHELIAL ADHESIN

; NUMBER OF SEQUENCES: 14

; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: CHRISTENSEN, O'CONNOR, JOHNSON AND KINDNESSPLLC
 ; STREET: SUITE 2800, 1420 FIFTH AVENUE
 ; CITY: SEATTLE
 ; STATE: WASHINGTON
 ; COUNTRY: USA
 ; ZIP: WA 98101
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IEM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US95/06994
 ; FILING DATE: 07-JUN-95
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/265,714
 ; FILING DATE: 24-JUN-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: BRODERICK, THOMAS F
 ; REGISTRATION NUMBER: 31,332
 ; REFERENCE/DOCKET NUMBER: CHOR-18591
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 682 8100
 ; TELEFAX: (206) 224 0779
 ; INFORMATION FOR SEQ ID NO: 5:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 696 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; PCT-US95-06994-5
 ;
 ; Alignment Scores:
 ; Pred. No.: 0 Length: 696
 ; Score: 3638.00 Matches: 696
 ; Percent Similarity: 100.00% Conservative: 0
 ; Best Local Similarity: 100.00% Mismatches: 0
 ; Query Match: 97.48% Indels: 0
 ; DB: 5 Gaps: 0
 ;
 ; US-10-625-972-4 (1-2091) x PCT-US95-06994-5 (1-696)
 ;
 ; QY 1 ATCCGAATAACCACTCTGCGTTCCTCGTAGTCATCCCTGCTCGGATTTTCAGCCAGCAGC 60
 ; Db 1 MetArgIleThrThrLeuAlaSerValValIleProCysLeuGlyPheSerAlaSerSer 20
 ; QY 61 ATAGCTGCTGCAGAGGATGTGATGTTCTCGGCATCCGCTATGAGAAAGCTGACT 120
 ; Db 21 IleAlaAlaGluAspValMetIleValSerAlaSerGlyTyrGluLysLysLeuThr 40
 ; QY 121 AACCCAGCCCGCAGTGTCTGTTCTGATTAGCCAGGAGGAATTCAGTCCAGCCAGTACCAC 180
 ; Db 41 AsnAlaAlaSerValSerValIleSerGlnGluLeuGlnSerSerGlnTyrHis 60
 ; QY 181 GATCTGGCGGAGCTCTGAGATCAGTAGGGGTGTGGATGTTGAAAGTGGTACGGGTAAA 240
 ; Db 61 AspLeuAlaGluAlaLeuArgSerValGluGlyValAspValGluSerGlyThrGlyLys 80
 ; QY 241 ACCGGAGGGCTGGAATCAGCATCCGAGGAATGCCAGCAGTTCACGCTGATCTGATT 300
 ; Db 81 ThrGlyGlyLeuGluIleSerIleArgGlyMetProAlaSerTyrThrIleIleIle 100
 ; QY 301 GATGTTGTTGTCAGGGCGGAAGCAGTGAAGTATCCCAACGGTTTTTCTGCCATGAAT 360
 ; Db 101 AspGlyValArgGlnGlySerSerAspValThrProAsnGlyPheSerAlaMetAsn 120
 ; QY 361 ACCGGTTTCATGCCCTCTGCGCCATTTGAGCGTATTGAGGTTATCAGGGGCGCGATG 420
 ; Db 121 ThrGlyPheMetProLeuAlaAlaIleGluArgIleGluValIleArgGlyProMet 140
 ; QY 421 TCCACATGTATGGCTCTGATGCGATGGCGGTGTGGTGAATATCATTTACCAGAAAGAAT 480

Db	141	SerThrLeuTyrGlySerAspAlaMetGlyGlyValValAsnIleIleThrArgLysAsn	160
Qy	481	CGACAGAAATGGCTCTCTTCGTCAATCCAGCGGTGAATCTGCAGGAACAATAATGG	540
Db	161	AlaAepLysTrpLeuSerSerValAsnAlaGlyLeuAsnLeuGlnGluserAsnLysTrp	180
Qy	541	GTAACACAGCACCAGTTAAATTCTGGAGCAGTGCTCCCTTGTGGATGATTCTGTCCAGC	600
Db	181	GlyAsnSerSerGlnPheAsnPhetripserGlyProleuValaspaspservalSer	200
Qy	601	CTGACAGGTACCGGTAGCACACAACAGCGTCAGGGTTTCATCGGTCACATCACTAGCGCAT	660
Db	201	LeuGlnValArgGlySerThrGlnArgGlnGlySerValThrSerLeuSerAsp	220
Qy	661	ACAGACAGCAGCGGTATTCCTATCCACAGGATCAGAGAATATATATCTGTGTGCACGT	720
Db	221	ThrAlaGlyThrArgIleProtyrProthrGluserGlnAsnTyAsnLeuGlyAlaArg	240
Qy	721	CTTGACTGGAAGCGCTCGAGCAGAGATGTCTCTGTTTGATFGATACACCCCGGCAG	780
Db	241	LeuAspTrpLysAlaSerGluGlnAspValLeutrpPheAspMetAspThrThrArgGln	260
Qy	781	CGTTATGATAACCGGATGGCAACTGGGAGTCTGACGGGGGATATGACCGGACCGCTG	840
Db	261	ArgTyrAspAsnArgAspGlyGlnneuGlySerLeuThrGlyGlyTyrAspArgThrLeu	280
Qy	841	CGCTATGACGGAACAAAATTTTCAGCTGGCTATGATCATCTTTTCACCTTCGGAACATGG	900
Db	281	ArgTyrGluArgAsnLysIleSerAlaGlyTyrAspHisThrPheThrPheGlyThrTrp	300
Qy	901	AAATCGTATCTGAATCGAACAGACAGAAATAAGTCTGTGACTTGTACGAGTGTA	960
Db	301	LysSerTyrLeuAsnTrpAsnGluThrGluAsnLysGlyArgGluLeuValArgSerVal	320
Qy	961	CTGAACGGCAGAAATGGGGGCTCCCGTTCAGCCGCCGGAGCTTAAGGAATCAAACCTT	1020
Db	321	LeuLysArgAspLysTrpGlyLeuAlaGlyGlnProArgGluLeuLysGluSerAsnLeu	340
Qy	1021	ATCCTGAATTCATTACTGTTACCCCTCTGGAGAAATCTATCTGTGTACGGTGGGGGC	1080
Db	341	IleLeuAsnSerLeuLeuLeuThrProLeuGlyGluSerHisLeuValThrValGlyGly	360
Qy	1081	GAGTTTCAGAGTCTCCATGAAGACGAGTGTCTTCGACGACACAGGTGAACATTC	1140
Db	361	GluPheGlnSerSerSerMetLysAspGlyValValLeuAlaSerThrGlyGluThrPhe	380
Qy	1141	CGGCAGAAAAAGTGTGTGGTATTTGCTGAGTAGTGATGCGATCTCACGGATGCGCTGCG	1200
Db	381	ArgGlnLysSerTrpSerValPheAlaGluAspGluTrpHisLeuThrAspAlaLeuAla	400
Qy	1201	CTGACTCGGGCAGCGCTATGNACATCATGACATTCGGGGACATTCAGTCCGCGT	1260
Db	401	LeuThrAlaGlySerArgTyrGluHisHisGluGlnPheGlyGlyHisPheSerProarg	420
Qy	1261	GCATATCTGCTCGGATGTGGCAGATGCTCGACGCTGAAAGCGGCTGCACCAACGGA	1320
Db	421	AlaTyrLeuValTrpaspValAlaAspAlaTrpThrLeuLysGlyGlyValThrThrGly	440
Qy	1321	TATAAGGCAACCAAGATGGGGAGCTACATAAAGGATTAGTGTGTCTCGGGCAGGGA	1380
Db	441	TyrLysAlaProArgMetGlyGlnLeuHisLysGlyIleSerGlyValSerGlyGlnGly	460
Qy	1381	AAACAARAATCTACTTGTATCCCCCGACCTCAAGCCCGAAGAGAGCGGTTCAGTTATCAGCT	1440
Db	461	LysThrAsnLeuLeuGlyAsnProAspLeuLysProGluGluSerValSerTyrGluAla	480
Qy	1441	GGGGTGATTACGATAACCCCGCGGTCTGAATGCOAATGTGCAGGTTTTATGACTGAC	1500
Db	481	GlyValTyrTyrAspAsnProAlaGlyLeuAsnAlaAsnValThrGlyPheMetThrAsp	500
Qy	1501	TTCTCCACACAGATGTCTCTATTCCATAAATGATAACCAATAGCTATGTAAACAGC	1560

501	PheSerAsnLysIleValSerTyrSerIleAsenAspAsnThrAsnSerTyrValasnSer	520
1561	GGAAAGCCCGGTTGCACGCTGTGGAAATTTGCCGGCACATTCCCGCTGTGGTCAGAGAT	1620
521	GlyLysIlaIargLeuHisGlyValGluPheAlaGlyThrLeuProLeuTrpSerGluAsp	540
1621	GTCACGCTGTCACTGAATATACACTCGACCCGACGAGTGAACACGTCATCGTATACAAA	1680
541	ValThrLeuSerLeuAsnTyrThrTrpThrArgSerGluGlnHArgaspGlyAspAsnLys	560
1681	GGTSCGGCGGTGAGTTATATACCCCTGAACACATCGTGAATGCGAAATCGAATCGGAGATC	1740
561	GlyAlaProLeuSerTyrThrProGluHisMetValasnAlaLysLeuAsnTrpGlnIle	580
1741	ACCGAAGAGGTGGCATCATGCTGGTGGTGGTCCCGTATCCGGGNAACACACCGTTTCACC	1800
581	ThrGluGluValAlaSerTrpLeuGlyAlaArgTyrArgGlyLysThrProArgPheThr	600
1801	CAGAAATATTCGTCACATGACGCTGTACAGAAAGATGTATCATGAGAAAGAGAGATAC	1860
601	GlnAsnTyrSerSerLeuSerAlaValGlnLysValTyrAspGluLysGlyGluTyr	620
1861	CTGAAGCCCTGACCGGTGGATGCGAGTCTCTCGTGAAGATGACGATGCCCTGACG	1920
621	LeuLysAlaTrpThrValValAspAlaGlyLeuSerTrpLysVetThrAspAlaLeuThr	640
1921	CTGAATGCTGGGTGAATAACCTGCTCAACAAAGATTACAGTCACGTCGCTGACCTGTACAGT	1980
641	LeuAsnAlaAlaValAsnAsnLeuLeuAsnLysAspTyrSerAspValSerLeuTyrSer	660
1981	GCCGTGAAGATGACGCTGTATGCCGGTGATTCTTCACAGCGGATCATCAACACAGGA	2040
661	AlaGlyLysSerThrLeuTyrAlaGlyAspTyrPheGlnThrGlySerSerThrThrGly	680
2041	TATGTGATACCTGACGGAATTAATCTGATGCTGCTGCACTATCAGTTC	2088
681	TyrValIleProGluArgAsnTyrTrpMetSerLeuAsnLysTrpGlnPhe	696

RESULT 4

Sequence 8, Application PC/TUS9506994
GENERAL INFORMATION:
APPLICANT: Children's Hospital & Medical Center
APPLICANT: University of Washington
APPLICANT: Washington State University Research Foundation
APPLICANT: TARR, PHILLIP I
APPLICANT: BILGE, SIMA S
APPLICANT: BESSER, THOMAS E
APPLICANT: VARY JR, JAMES C
TITLE OF INVENTION: ESCHERICHIA COLI O157:H7 EPITHELIAL ADHESIN
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: CHRISTENSEN, O'CONNOR, JOHNSON AND KINDNESSPLLC
STREET: SUITE 2800, 1420 FIFTH AVENUE
CITY: SEATTLE
STATE: WASHINGTON
COUNTRY: USA
ZIP: WA 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/06994
FILING DATE: 07-JUN-95
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/265,714
FILING DATE: 24-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: BRODERICK, THOMAS F
REGISTRATION NUMBER: 31,332

REFERENCE/DOCKET NUMBER: CHOR-18591

TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 682 8100
 TELEFAX: (206) 224 0779
 INFORMATION FOR SEQ ID NO: 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 703 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 DESCRIPTION: E. coli O157:H7 adhesin amino acid sequence,
 wherein "Xaa" residues represent gaps
 DESCRIPTION: introduced to facilitate best alignment with
 DESCRIPTION: SEQ ID NO:9.
 HYPOTHETICAL: NO
 ORIGINAL SOURCE:
 ORGANISM: Escherichia coli O157:H7
 STRAIN: 86-24 NALR
 PCT-US95-06994-8

Alignment Scores:

Pred. No.: 0 Length: 703
 Score: 394.50 Matches: 695
 Percent Similarity: 98.86% Conservative: 0
 Best Local Similarity: 98.86% Mismatches: 1
 Query Match: 96.32% Indels: 7
 DB: 5 Gaps: 3

US-10-625-972-4 (1-2091) x PCT-US95-06994-8 (1-703)

QY	1	ATCGGAATAACCACTCTGGCTCCGTCAGTCATCCCTGTCGGATTTTCACCGCAGCAGC	60
Db	1	MetArgileThrThrLeuAlaSerValValIleProCysLeuGlyPheSerAlaSerSer	20
QY	61	ATA-----CCTCTGTCGAGGATGATGATGCTCTCGGCATCCGGCTAT	105
Db	21	Ile-----AlaAlaAlaGluAspValMetIleValSerAlaSerGlyTyr	40
QY	106	GAGAAAGACTGACTAACCGCCGCGCTGCTGTCGATTCGTCAGGAGAAATTCGAG	165
Db	41	GluLysLysLeuThrAsnAlaAlaAlaSerValSerValIleSerGlnGluLeuGln	60
QY	166	TCAGCGCATACAGATCTGCGGAGGCTCTGACATCAGTAGAGGCTGCGATTCGAA	225
Db	61	SerSerGlnTyrHisAspLeuAlaGluAlaLeuArgSerValGluGlyValAspValGlu	80
QY	226	ATGCTAGCGGTAAACCGGAGGCTGGAATACAGATCCGAGGAATCCAGCCAGTTAC	285
Db	81	SerGlyThrGlyLysThrGlyGlyLeuGluIleSerIleArgGlyMetProAlaSerTyr	100
QY	286	ACGCTGATCTATGATGGTGTCTGTCAGGCGGAGACAGTACGTGACTCCCAAGGT	345
Db	101	ThrLeuIleLeuLeuAspGlyValArgGlnGlySerSerAspValThrProAsnGly	120
QY	346	TTTTCTGTCATGAATACCGGTTCTATGCCCTCTGCGCCGCTTGGCGTATTGAGGTT	405
Db	121	PheSerAlaMetAsnThrGlyPheMetProLeuAlaAlaIleGluArgIleGluVal	140
QY	406	ATCAGGGGCGCATGTCACACTGTATGGCTCTGATGCGATGGCGGTGGTGAATATC	465
Db	141	IleArgGlyProMetSerThrLeuTyrGlySerAspAlaMetGlyValValAsnIle	160
QY	466	ATTACAGAAAGATGCAGACAAATGGCTCTCTCCGTCATGTCAGGCGTGAATCTCAG	525
Db	161	IleThrArgLysAsnAlaAspLysTrpLeuSerSerValAsnAlaGlyLeuAsnLeuGln	180
QY	526	GAAAGCAACAAATGGGTAAACAGCAGCAGTAAATTTCTGGAGCAGTGGTCCCTTGTG	585
Db	181	GluSerAsnLysTrpGlyAsnSerSerGlnPheAsnThrTrpSerSerGlyProLeuVal	200
QY	586	GATGATCTGTGCTCAGCTCAGCGGTAGCACACACAGCTCAGGCTTCATCGGTC	645
Db	201	AspAspSerValSerLeuGlnValArgGlySerThrGlnArgGlnGlySerSerVal	220

QY	646	ACATCACTAGCGGATACAGCAGGACCGGTATTCTCTTATCCACGGAGTCACAGAATTAT	705
Db	221	ThrSerLeuSerAspThrAlaGlyThrArgIleProTyrProThrGluSerGlnAsnTyr	240
QY	706	AATCTTGGTCACGCTTGTGACTGGAGCGCTCGGAGCAGGATGCTCTCGTTGATATG	765
Db	241	AsnLeuGlyAlaArgLeuAspTrpLysAlaSerGlnGlnAspValLeuTrpPheAspMet	260
QY	766	GATACACCCGGCAGCGCTTATGATAACCGGATGCGCACTGGGGAGTCTGACGGGGGA	825
Db	261	AspThrThrArgGlnArgTyrAspAsnArgAspGlyGlnLeuGlySerLeuThrGlyGly	280
QY	826	TATGACCGGACCTGCGCTATGACGCAACAAATTCAGCTGGCTATGATCATCTTTC	885
Db	281	TyrAspArgThrLeuArgTyrGluArgAsnLysIleSerAlaGlyTyrAspGlyThrPhe	300
QY	886	ACCTTCGGAACATCGAAATCGTATCTGAACCTGGAACGAGACAGAAAAATAAGTCTGAG	945
Db	301	ThrPheGlyThrTrpLysSerTyrLeuAsnTrpAsnGluThrGluAsnLysGlyArgGlu	320
QY	946	CTTGACCGCAGTGTACTGAAGCGCAAAATGGGGCTTCGCGTCAGCCGGGAGCTT	1005
Db	321	LeuValArgSerValLeuLysArgAspLysTrpGlyLeuAlaGlyGlnProArgGluLeu	340
QY	1006	AAGGAATCGAACCTTATCTGAATTCATTACTCTTACCCCTCTGGGAGAAATCTCATCG	1065
Db	341	LysGluSerAsnLeuIleLeuAsnSerLeuLeuLeuThrProLeuGlyGluSerHisLeu	360
QY	1066	GTTACGGTGGGGCGAGTTTCAGAGCTCTCATGAAAGACGGAGTTCCTCTTCCAGC	1125
Db	361	ValThrValGlyGlyGluPheGlnSerSerMetLysAspGlyValValLeuAlaSer	380
QY	1126	ACAGGTGAAACTTTCGGCAGCAAAAGC---TGGTGGTATTTCGAGAGTGGGAT	1182
Db	381	ThrGlyGluThrPheArgGlnLysSer---TrpSerValPheAlaGluAspGluTrpHis	400
QY	1183	CTCACGGATGCATTCGCTGCTGCTGCGGCGCAGCTATGAACATCATGAGCAATTCGGG	1242
Db	401	LeuThrAspAlaLeuAlaLeuThrAlaGlySerArgTyrGluHisGluGlnPheGly	420
QY	1243	GGACACTCAGTCCGCGTGCATATCTGGTCTGGAGTGGCAGATGCCCTGGACGTGAA	1302
Db	421	GlyHisPheSerProArgAlaTyrLeuValTrpAspValAlaAspAlaTrpThrLeuLys	440
QY	1303	GGCGGTGTGACACACGGGATATAAGGCACCCAGAAATGGGCGAGCTACATAAAGGGATTAGT	1362
Db	441	GlyGlyValThrThrGlyTyrLysAlaProArgMetGlyGlnLeuHisLysGlyLysSer	460
QY	1363	GGTGTGTCCGGCAGGAAAAACAATCTACTTGGTAACCCCGACTGAAGCCGGAAGAG	1422
Db	461	GlyValSerGlyGlnGlyLysThrAsnLeuLeuGlyAsnProAspLeuLysProGluGlu	480
QY	1423	AGCGTCAGTTATGAGCTGGGTGATTACGATAACCCCGCGTCTGAATCCCAATGTC	1482
Db	481	SerValSerTyrGluAlaGlyValTyrAspAsnProAlaGlyLeuAsnAlaAsnVal	500
QY	1483	ACAGGTTTTATGACTGACTTCTCCAAACAAGATT---GTCTCTTATTCCATAAATGATAAC	1539
Db	501	ThrGlyPheMetThrAspPheSerAsnLysIle---ValSerTyrSerIleAsnAspAsn	520
QY	1540	ACCAATAGCTATGTAACACAGCGGAAGCCCGGTTCACGGTGTGCAATTTGCCGGACA	1599
Db	521	ThrAsnSerTyrValAsnSerGlyLysAlaArgLeuHisGlyValGluPheAlaGlyThr	540
QY	1600	TTCCCGCTGTGTGTCAGAGGATGTCACGCTGTCACCTGCAATTAACCTGACCCGAGTCAA	1659
Db	541	LeuProLeuTrpSerGluAspValThrLeuSerLeuAsnTyrThrTrpThrArgSerGlu	560
QY	1660	CAACCTGATGGTATAACAAAGGTCCGCGCTGAGTTATACCCCTGACACACATGGTGAAT	1719
Db	561	GlnArgAspGlyAspAsnLysGlyAlaProLeuSerTyrThrProGluHisMetValAsn	580

Db 261 AspValLeuTrpPheAspMetAspThrThrArgGlnArgTyrAspAsnArgAspGlyGln 280
QY 805 CTGGGAGTCTGACGGGGGATATGACCGGACCTCGCTGATGAGCGAACAATAATTTCA 864
Db 281 LeuGlySerLeuThrGlyGlyTyrAspArgThrLeuArgTyrGluArgAsnLysIleSer 300
QY 865 GCTGGCTATCATCATCTTCCCTTCGGAACATGGAATTCCTATCTGAACCTGGACGAG 924
Db 301 AlaGlyTyrAspGlyThrPheThrPheGlyThrTrpLysSerTyrLeuAsnTrpAsnGlu 320
QY 925 ACAGAAATAAAGCTCGTACGCTTGTACGACAGTACTGAAGCGGCACAAATGGGGCTT 984
Db 321 ThrGluAsnLysGlyArgGluLeuValArgSerValLeuLysArgAspLysTrpGlyLeu 340
QY 985 GCCGTCACCGCGGAGCTTAAGGAATCGAACCTTATCTCAATTCATTCTGCTTACC 1044
Db 341 AlaGlyGlnProArgGluLeuLysGluSerAsnLeuIleLeuAsnSerLeuLeuLeuThr 360
QY 1045 CCTCTCGGAGAACTCTCATCTGTTACGCTGGGGGCGAGTTTCAGAGCTCGTCCATGAAA 1104
Db 361 ProLeuGlyGluSerHisLeuValThrValGlyGlyGluPheGlnSerSerMetLys 380
QY 1105 GACGGAGTGTCTTCCAGCAGCAGGTGAAACTTTCGGCAGAAAAGCTGTGCTGTTT 1164
Db 381 AspGlyValValLeuAlaSerThrGlyGluThrPheArgGlnLysSerTrpSerValPhe 400
QY 1165 GCTGAGGATGAGTGGCATCTCAGGATGACCTTGGCTGCTGCGGCGGAGCGCTATGAA 1224
Db 401 AlaGluAspGluTrpHisLeuThrAspAlaLeuAlaLeuThrAlaGlySerArgTyrGlu 420
QY 1225 CATCATGAGCAATTCGGGGACACTTCAGTCCGCTGCTATCTGCTGGGATGTGCA 1284
Db 421 HisHisGlnGlnPheGlyGlyHisPheSerProArgAlaTyrLeuValTrpAspValAla 440
QY 1285 GATGCTGAGCGCTGAAAGCGGTGTGACCGGATATAGGCACCCAGATGGGCGAG 1344
Db 441 AspAlaTrpThrLeuLysGlyValThrThrGlyTyrLysAlaProArgValGlyGln 460
QY 1345 CTACATAAAGGATTAGTGTGTGTCGGGCGAGGAAACAAATCTACTTGGTAACCCC 1404
Db 461 LeuHisLysGlyIleSerGlyValSerGlyGlnGlyLysThrAsnLeuLeuGlyAsnPro 480
QY 1405 GACCTGAACCCGAGAGCGCTCAGTTATGAGCTGGGTGTATTAGATAACCCGCC 1464
Db 481 AspLeuLysProGluGluSerValSerTyrGluAlaGlyValTyrTyrAspAsnProAla 500
QY 1465 GGTCTGAATGCCAATGTACAGGTTTATGCTGACTTCTCCAAACAAGATT-----GTC 1518
Db 501 GlyLeuAsnAlaAsnValThrGlyPheMetThrAspPheSerAsnLysIle*****Val 520
QY 1519 TCTTATTCATAAATGATACACCAATAGCTATGTAACACAGCGGAAAGGCC----- 1569
Db 521 SerTyrSerIleAsnAspAsnThrAsnSerTyrValAsnSerGlyLysAla***** 540
QY 1570 CGTTTCACGGGTGGAAATTTCCCGGC-----ACATTCGCGCTGTGTCA 1614
Db 541 ArgLeuHisGlyValGluPheAlaGly*****ThrLeuProLeuTrpSer 560
QY 1615 GAGGATGTCACCTCTCATTAATACCTGATACCTGACACATGTTGTAATCGGAACCTGAG 1674
Db 561 GluAspValThrLeuSerLeuAsnTyrThrTrpThrArgSerGluGlnArgAspGlyAsp 580
QY 1675 AACAAAGTGGCCCTCAGTTATACCTGATACCTGATGATGTAATCGGAACCTGAGTGG 1734
Db 581 AsnLysGlyAlaProLeuSerTyrThrProGluHisMetValAsnAlaLysLeuAsnTrp 600
QY 1735 CAGATCACCGAAGAGTGGCATCATGCTGGGTGCCGTTATCGCGGAAACACCAACGT 1794
Db 601 GlnIleThrGluGluValAlaSerTrpLeuGlyAlaArgTyrArgGlyLysThrProArg 620
QY 1795 TTCACCGAGATTATCTGCTGAGCGCTGTACAGAAAGTGTATGATGAGAAAGGA 1854

Db 621 PheThrGlnAsnTyrSerSerLeuSerAlaValGlnLysLysValTyrAspGluLysGly 640
QY 1855 GAATACCTGAAAGCTCGACGGTGGATGCGAGTCTGTCTGTTGGAAGATGACCGATGCC 1914
Db 641 GluTyrLeuLysAlaTrpThrValValAspAlaGlyLeuSerTrpLysMetThrAspAla 660
QY 1915 CTGACGCTGAATGCTCGGTGAATACCTGCTCAACAAGGATTACAGTACGCTGACGCTG 1974
Db 661 LeuThrLeuAsnAlaAlaValAsnLeuLeuAsnLysAspTyrSerAspValSerLeu 680
QY 1975 TACAGTCCCGTAAAGTACGCTGTATGCCGCTGATTACTCCAGACGGGATCATCAACA 2034
Db 681 TyrSerAlaGlyLysSerThrLeuTyrAlaGlyAspTyrPheGlnThrGlySerSerThr 700
QY 2035 ACAGGATATGATACCTGAGCGGAATTAATCTGAGTCTCGCTGAACTATCATGTTTC 2088
Db 701 ThrGlyTyrValIleProGluArgAsnTyrTrpMetSerLeuAsnTyrGlnPhe 718
RESULT 6
PCT-US95-06994-7
; Sequence 7, Application PC/TUS9506994
; GENERAL INFORMATION:
; APPLICANT: Children's Hospital & Medical Center
; APPLICANT: University of Washington
; APPLICANT: Washington State University Research Foundation
; APPLICANT: TARR, PHILLIP I
; APPLICANT: BILGE, SIMA S
; APPLICANT: BESSER, THOMAS E
; APPLICANT: VARY JR, JAMES C
; TITLE OF INVENTION: ESCHERICHIA COLI O157:H7 EPITHELIAL ADHESIN
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESS: CHRISTENSEN, O'CONNOR, JOHNSON AND KINDNESS PLLC
; STREET: SUITE 2800, 1420 FIFTH AVENUE
; CITY: SEATTLE
; STATE: WASHINGTON
; COUNTRY: USA
; ZIP: WA 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06994
; FILING DATE: 07-JUN-95
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,714
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BRODERICK, THOMAS F
; REGISTRATION NUMBER: 31,332
; REFERENCE/DOCKET NUMBER: CHOR-18591
; TELEPHONE: (206) 582 8100
; TELEFAX: (206) 224 0779
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 718 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; DESCRIPTION: Vibrio cholerae Irga amino acid sequence,
; DESCRIPTION: wherein xaa" residues represent gaps
; DESCRIPTION: introduced to facilitate best alignment with
; DESCRIPTION: SEQ ID NO:6.
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: VIBRIO CHOLERAE
; PCT-US95-06994-7
Alignment Scores:

Pred. No.:	1.9e-91	Length:	718
Score:	1089.50	Matches:	254
Percent Similarity:	51.53%	Conservative:	117
Best Local Similarity:	35.28%	Mismatches:	294
Query Match:	29.19%	Indels:	55
DB:	5	Gaps:	13

US-10-625-972-4 (1-2091) x PCT-US95-06994-7 (1-718)			
QY	40	CTCGGATTTTCAGCCAGCAGCATAGCTGCTGCAGAGGAT-----GTG	81
DB	17	LeuMetPheSerAlaSer*****AlaPheAlaGlnAspAlaThrLysThrAspGluThr	36
QY	82	ATGATTTGCTCGGATCCGGCTATGAGAAAAGCTGACTAACGAGCGCGCAGTGTCT	141
DB	37	MetValValThrAlaAlaGlyTyAlaGlnValIleGlnAsnAlaProAlaSerIleSer	56
QY	142	GTGATTACCCAGGAGGAATTCAGTCCAGCCAGTACCACCATCTCGGGGAGGCTCTGAGA	201
DB	57	ValIleSerArgGluAspLeuGluSerArgTyTyArgAspValThrAspAlaLeuLys	76
QY	202	TCAGTAGAGGTGGATGTTGAAGTGGTACGGGTAAACCGGAGGCTGGAAATCAGC	261
DB	77	SerValProGlyValThrVal*****ThrGlyGlyAspThrThrAspIleSer	96
QY	262	ATCCGAGGAATGCCAGCCAGTACAGCTGATCTGATGATGTTGTCAGGCGCGA	321
DB	97	IleArgGlyMetGlySerAsnTyThrLeuIleLeuValAspGlyLysArgGln---Thr	115
QY	322	AGCAGTACGCTGACTCCCAACGGTTTTTCT---GCCATGAATACCGGTTTCATCCCT	378
DB	116	SerArgGlnThrArgProAsnSerAspGlyProGlyIleGluGlnGlyTrpLeuPro	135
QY	379	CTGCGCCGCAATGACGTAATGAGTTATCAGGGGCGGATCTCCACACTGTATGGCTCT	438
DB	136	LeuGlnAlaIleGluArgIleGluValIleArgGlyProMetSerThrLeuTyGlySer	155
QY	439	GATGGATGGCGGTGGTGGTGAATATCATATACCAAGAAATGCAGACAAATGCTCTCT	498
DB	156	AspAlaIleGlyValIleAsnIleIleThrArgLysAspGlnGlnTrpSerGly	175
QY	499	TCCGTCAATGCGGCTGAATCTGCAGGAAGCAAAATGGGTAAACAGCAGCGAGTTT	558
DB	176	AsnValGlnLeuSerThrValValGlnGluAsnArgAlaSerGlyAspGluGlnSerAla	195
QY	559	AAATTTCTGAGCAGNGTCCCTGCTGGATGATCTGTGACGCTGCAGCTACGGGTAGC	618
DB	196	AsnPhePheValThrGlyProLeuSerAsp***AlaLeuSerLeuGlnValTyGlyGln	215
QY	619	ACACAACAGCGTCAGGGTTTCATCGGTCACTACATCACTGACCGATACAGCAGCGGAT	678
DB	216	ThrThrGlnArgAspGluAspGluIleGluHis-----	226
QY	679	CCTTATCCACCGAGTCACAGAAATATATCTTGTGTCAGCTCTGATCGGAGCGTCG	738
DB	227	GlyTyGlyAspLysSerLeuArg---SerLeuThrSerLysLeuAsnTyGlnLeuAsn	245
QY	739	GAGCAGGATGCTCTGCTGTTTATGATATGATACCCCGGCGGCTTATGATAACCGGAT	798
DB	246	ProAspHisGlnLeuGlnLeuGluAlaGlyValSerAlaGlnAspArgGluAsnVal	265
QY	799	GGGCAACTGGGAGTCTAGCGGGGGATATGACCGGACCGCTG-----	840
DB	266	GlyLysSerAlaGlnSerSerGlyCys*****ArgGlyThrCysSer	285
QY	841	-----CGCTATGAGGGAACAAATTTTCAGCTGGCTGATGATCATCTTTCACC	888
DB	286	AsnThrAspAsnGlnTyArgArgAsnHisValAlaValSerHisGln-----	301
QY	889	TTCCGAACTGGAATA-----TCGTATCTGAACTGGAACGAGACA	927
DB	302	---GlyAspTrpGlnGlyVal***GlyGlnSerAspThrTyLeuGlnTyTrpGluAsn	320

QY	928	GAAATTAAGGTCGTGAGCTTGTACGAGTGTACTGAAGCGGCACAAATGGGGCTTGCC	987
DB	321	ThrAsnLysSerArgGlu*****	340
QY	988	GGTCAGCCCGGAGCTTAAGGAATCGAACCTTATCTCGAATTCATCTGCTTACCCT	1047
DB	341	*****MetSerIleAspAsnThrValPheLysSerThrLeuValAlaPro	360
QY	1048	CTGGGAGATCTCATCTGTTTACGTTGGGGGCGGAGTTTCAGAGCTCGTCATGAAGAC	1107
DB	361	IleGlyGlu---HisMetLeuSerPheGlyValGluGlyLysHisLeuSerLeuGluAsp	379
QY	1108	GGAGTTGTCTTCCAGCAGCAGGTGAACCT---TTCGGCAGAAAAGCTGTCGGTATT	1164
DB	380	LysThrSerAsnLysIleSerSerArgThrHisIleSerAsnThrGlnTrpAlaGlyPhe	399
QY	1165	GCTGAGGATGAGTGGCATCTCAGGATGCATTCGCTGCTGCTCGCGGAGCGCTATGAA	1224
DB	400	IleGluAspGluTrpAlaLeuAlaGluGlnPheArgLeuThrPheGlyGlyArgLeuAsp	419
QY	1225	CATCATGAGCAATTCGGGGGACATTCAGTCCGCGTGCATATCTGCTCTGGGATGGCA	1284
DB	420	HisAspLysAsnTyGlySerHisPheSerProArgValTyGlyValTrpAsnLeuAsp	439
QY	1285	GATGCTGACGCTGAAGCGGTGTGACACGGGATATAAGCACCAGATGGGCGAG	1344
DB	440	ProLeuTrpThrValLysGlyGlyValSerThrGlyPheArgAlaProGlnLeuArgGlu	459
QY	1345	CTACATAAAGGATTAGTGTGTCGGCGGAGGAAAACAAATCTACTTGGTAAACCC	1404
DB	460	ValThrProAspTrpGlyGlnValSerGlyGly*****AsnIleTyGlyAsnPro	479
QY	1405	GACTGAAGCCGGAAGAGCGCTCAGTTATGAGGTGGGTGTTATACGATAACCCCGCC	1464
DB	480	AspLeuGlnProGluThrSerIleAsnLysGluLeuSerLeuMetTySerThrGlySer	499
QY	1465	GGTCTGAATGCCAATCTCAGCGTTTATGACTGCTCTCCAAACAGATTTGCTCTTAT	1524
DB	500	GlyLeuAlaAlaSerLeuThrAlaPheHisAsnAspPheLysAspLysIleThrArgVal	519
QY	1525	TCCATAAATGATAAC-----ACCAATAGCTAT	1551
DB	520	AlaCysProAlaAsnIleCysThrAlaGlyProAsnGlnTrpGlyAlaThrProThrTy	539
QY	1552	---GTAACACCGGAAGAGCCCGGTGCGAGTGTGGAATTTCCCGCAGATTCGCGCTG	1608
DB	540	ArgValAsnIleAspGluAlaGluThrTyGlyAlaGluAlaThrLeuSerLeuProIle	559
QY	1609	TGCTCAGAGGATCTCAGCTGTCTCAGTGAATACACCTGACCCGAAAGTGAACACGTAT	1668
DB	560	---ThrGluSerValGluLeuSerSerTyThrTyThrHisSerGluGlnLysSer	578
QY	1669	GGTGATAACAAAGGTGGCGCTGAGTTATACCCCTGAACACATGTTGAATGCGAACTG	1728
DB	579	GlyAsnPheAlaGlyArgProLeuLeuGlnLeuProLysHisLeuPheAsnAlaAsnLeu	598
QY	1729	AACCTGCAGATCCCGNAGAGTGGCATGCTGCTGCTCAACAGGATTAACAGTGC	1788
DB	599	SerTrpGlnThrThrArgLeuAsnSerTrpAlaAsnLeuAsnTyArgGlyLys***	618
QY	1789	CCAGTTTCCACCCAGATTAATTCGTCTCAGCGCTGTACAGAGAAAGTGTATGATGAG	1848
DB	619	*****GluMetGlnProGluGlyGlyAlaSerAsnAspPheIleAlaPro*****	638
QY	1849	AAAGGAGAAATACCTGAAGCTGACGCTGCTGATGATGATGCTGCTGCTGGAAGATGAC	1908
DB	639	*****SerTyThrPheIleAspThrGlyValThrTyAlaLeuThr	658
QY	1909	GATCCCTGAGCTGATGCTGCGGTGAATACTGCTCAACAGGATTAACAGTGC	1968
DB	659	AspThrAlaThrIleLysAlaAlaValTyAsnLeuPheAspGlnGlu*****	678
QY	1969	AGCCTGTACAGTCCCGGTAAGAGTACGCTGTATGCCGGTGAATCTTCCAGACGGATCA	2028

Db	57	SerArgGluAspLeuGluSerArgTyrTyrArgAspValThrAspAlaLeuLysSerVal	76
Qy	208	GAGGGTGTGGATGTTGAAGTGGTACGGGTAAACCCGAGGCGTGGAAATCAGCATCCGA	267
Db	77	ProGlyValThrValThrGlyGlyAspThrThr-----AspIleSerIleArg	93
Qy	268	GGAATGCCAGCAGTTACACGCTGATACTGATGATGATGTTGTCAGCGCGGAAGCAGT	327
Db	94	GlyMetGlySerAsnTyrThrLeuIleLeuValAspGlyLysArgGln---ThrSerArg	112
Qy	328	GACGTGACTCCCAACGGTTTTCT---GCCATGAATACCGGGTTCATCCGCCCTCTGGCC	384
Db	113	GlnThrArgProAsnSerAspGlyProGlyIleGluGlnGlyTyrLeuProLeuGln	132
Qy	385	GCCATTGAGCGTATTGAGGTTCATCAGGGGCGCGATGTCACACTGTTATGGTCTCTGATCG	444
Db	133	AlaIleGluArgIleGluValIleArgGlyProMetSerThrLeuTyrGlySerAspAla	152
Qy	445	ATGGCGGTGTGGTAATATCATTTACCGAAAGAATCAGACAATGGCTCTCTTCGCT	504
Db	153	IleGlyGlyValIleAsnIleIleThrArgLysAspGlnGlnIntrpSerGlyAsnVal	172
Qy	505	AATCGAGGCTGAATCTCGACGAAGCAACAAATGGGTAAACAGCAGCCAGTTTAATTTC	564
Db	173	GlnLeuSerThrValValGlnGluAsnArgAlaSerGlyAspGluGlnSerAlaAsnPhe	192
Qy	565	TGGAGCAGTGTCCCTTGTGGATGATTTCTCAGCTCGCAGGTACGGGTAGCACAA	624
Db	193	PheValThrGlyProLeu---SerAspAlaLeuSerLeuGlnValTyrGlyInThrThr	211
Qy	625	CAGGTCAG-----GGTTCATCGGTACACATCACTGAGC	657
Db	212	GlnArgAspGluAspGluIleGluHisGlyTyrGlyAspLysSerLeuArgSerLeu---	230
Qy	658	GATACAGCAGCACCGCTATTCTTTATCCACCGAGTCACAGAATTATACTTTGGT---	714
Db	231	-----ThrSerLysLeuAsnTyrGlnLeuAsnPro	240
Qy	715	-----GCACGTCTGACTGGAAGCGCTCGGAGCAGGATGTGCTCTGGTTTCATATG	765
Db	241	AspHisGlnLeuGlnLeuGluAlaGlyValSerAlaGlnAsp-----	254
Qy	766	GATACCAACCGGACCGCTTATGATAACCGGATGGGCAACTGGGGAGCTGTGACCGGGGGA	825
Db	255	-----ArgGluAsnValGlyLysSerAlaGlnSerSerGlyCysArgGlyThr	271
Qy	826	TATGACCGGACC---CTGCGCTATGACGGAACAAATTTTCAGCTGCTGATGATCATCT	882
Db	272	CysSerAsnThrAspAsnGlnTyrArgArgAsnHisValAlaValSerHisGln-----	289
Qy	883	TTACCTTCGGAACATCGAAA-----TCGTATCTGAACTGGAAACGAG	924
Db	290	-----GlyAspTyrGlnGlyValGlyGlnSerAspThrTyrLeuGlnTyrGluGlu	306
Qy	925	ACAGAAATAAGTCTGTAGCTGTGTACGCGTGTACTGAAGCGCGACAAATGGGGGCTT	984
Db	307	AsnThrAsnLysSer-----	311
Qy	985	GCCGGTCAGCCGGGAGCTTAAGGAATCGAACCCTTATCTCGAATTCATTACTGCTTACC	1044
Db	312	-----ArgGluMetSerIleAspAsnThrValPheLysSerThrLeuValAla	327
Qy	1045	CCTCTGGGAATCTCATCTGTTTACGGTCGGGGGCGAGTTTCAGAGCTCGTCCATGAAA	1104
Db	328	ProIleGlyGlu---HisMetLeuSerPheGlyValGluGlyHisGluSerLeuGlu	346
Qy	1105	GACGGAGTTCTCTGCCACACACAGGTGAACCT---TTCCGGCGAGAAAACCTCGTCGTA	1161
Db	347	AspLysThrSerAsnLysIleSerSerArgThrHisIleSerAsnThrGlnTyrPalaGly	366
Qy	1162	TTTGCTGAGGATGAGTGGCATCTCACGGATGCATCTGCGCTGACTCGGGCAGCCGCTAT	1221


```

Db      367 PheIleGluAspGluTrpAlaLeuAlaGluGlnPheArgLeuThrPheGlyGlyArgLeu 386
Qy      1222 GAACATCATGAGCAATTCGGGGACACTTCAGTCGCGTCATATCTGCTCGGATGTG 1281
Db      387 AspHisAspLysAsnTyrGlySerHisPheSerProArgValTyrGlyValTrpAsnLeu 406
Qy      1282 GCAGATCCCTGGAGCTGAAAGCGGTGTACACCGGGATATAAGGACCCAGCAATGGGG 1341
Db      407 AspProLeuTrpThrValLysGlyValSerThrGlyPheArgAlaProGlnLeuArg 426
Qy      1342 CAGCTACATAAAGGATTAGTGTGTCTCCGGGAGGAAAAACAATCTACTTGTGTAAAC 1401
Db      427 GluValThrProAspTrpGlyGlnValSerGlyGly-----AsnIleTyrGlyAsn 444
Qy      1402 CCCGACTGAAGCCGGAAGAGAGCGTCAGTATACAGCTGGGTGTATTACGATAACCCC 1461
Db      445 ProAspLeuGlnProGluThrSerIleAsnLysGluLeuSerLeuMetTyrSerThrGly 464
Qy      1462 GCCGCTCTGAATGCAATGTACAGGTTTATGACTGACTTCTCCACAAGATTGTCTCT 1521
Db      465 SerGlyLeuAlaAlaSerLeuThrAlaPheHisAsnAspPheLysAspLysIleThrArg 484
Qy      1522 TATTCCTAAATGATAAC-----ACCAATAGC 1548
Db      485 ValAlaCysProAlaAsnIleCysThrAlaGlyProAsnGlnTrpGlyAlaThrProThr 504
Qy      1549 TAT---GTAACAGCGGAAGCGCGGTGCACGCTGTGGAATTTGCCGGCACATTGCCG 1605
Db      505 TyrArgValAsnIleAspGluAlaGluThrTyrGlyAlaGluAlaThrLeuSerLeuPro 524
Qy      1606 CTGTGGTCAGAGGATGTCACCTGTCACCTGAATATACCTGACCTGGACCGAAGTGAACAGT 1665
Db      525 Ile---ThrGluSerValGluLeuSerSerTyrThrTyrThrHisSerGluGlnLys 543
Qy      1666 GATGTGTATACAAAGGTGCCCGCTGAGTATATACCTGACATGCGTGAACATGGTGAATCGCAA 1725
Db      544 SerGlyAsnPheAlaGlyArgProLeuLeuGlnLeuProLysHisLeuPheAsnAlaAsn 563
Qy      1726 CTGAACCTGGCAGATCACCGAAGAGTGGCATGCTGGTGGTGGCGCTTATCGCGGAAA 1785
Db      564 LeuSerTrpGlnThrThrAspArgLeuAsnSerTrpAlaAsnLeuAsnTyrArgGlyLys 583
Qy      1786 ACACACGTTTACCACAGAAATTATTCGTCAGTACGCGCTGTACAGAAAGTGTATGAT 1845
Db      584 -----GluMetGlnProGluGlyGlyAla 591
Qy      1846 GAGAAAGGAGATACCTG---AAGCCCTGGAGGTGGTGGATGCGAGGTCTCTCTGGAAG 1902
Db      592 SerAsnAspAspPheIleAlaProSerTyrThrPheIleAspThrGlyValThrTyrAla 611
Qy      1903 ATGACGGATGCCCTGACCTGAATCTGCGGTGAATAACCTGCTCAACAGGATTACAGT 1962
Db      612 LeuThrAspThrAlaThrIleLysAlaAlaValTyrAsnLeuPheAspGlnGluValAsn 631
Qy      1963 GACGTGAGCCTGTACAGTCCCGTAAGAGTACGCTGTATGCGGTGATCTTCCAGACG 2022
Db      632 -----TyrAla---GluIyr----- 635
Qy      2023 GGATCATCAACACAGGATATGTATACCTGACGCAAAATTACTGATCTCGCTGAACAT 2082
Db      636 -----GlyTyrValGluAspGlyArgTyrTyrTriLeuGlyLeuAspIle 650
Qy      2083 CAGTTC 2088
Db      651 AlaPhe 652

```

RESULT 8

```

US-09-098-082-6
; Sequence 6, Application US/09098082
; Patent No. 6040421
; GENERAL INFORMATION:
; APPLICANT: Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
; TITLE OF INVENTION: Escherichia Coli 0157:H7 Epithelial Adhesin

```

```

; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
; STREET: 2800 Pacific First Center, 1420 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette-3.5 inch, 1.44MB storage
; COMPUTER: IBM PC compatible/Pentium II
; OPERATING SYSTEM: MS-Windows 95
; SOFTWARE: Word for Windows-6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/098,082
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/765,081
; FILING DATE: March 26, 1997
; APPLICATION NUMBER: PCT/US95/06994
; FILING DATE: June 7, 1995
; APPLICATION NUMBER: US 08/265,714
; FILING DATE: June 24, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Sheiness, Diana K.
; REGISTRATION NUMBER: 35,356
; REFERENCE/DOCKET NUMBER: CHOR-1-12402
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 1-206-682-8100; 1-206-224-0735 (direct)
; TELEFAX: 1-206-224-0779
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 652 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; DESCRIPTION: Vibrio cholerae IrgA amino acid sequence
; HYPOTHEICAL: NO
; ORGANISM: Vibrio Cholerae
; ORGANISM: Vibrio Cholerae
; US-09-098-082-6

```

```

Alignment Scores:
Pred. No.: 4,57e-88 Length: 652
Score: 1052.50 Matches: 258
Percent Similarity: 51.94% Conserved: 117
Best Local Similarity: 35.73% Mismatches: 222
Query Match: 28.20% Indels: 125
DB: 3 Gaps: 24

```

US-10-625-972-4 (1-2091) x US-09-098-082-6 (1-652)

```

Qy      40 CTGGATTTTCAGCCAGCAGCATAGCT-----GCTGCAGAGGTGTGATGATT 87
Db      17 LeuMetPheSerAlaSerAlaPheAlaGlnAspAlaThrLysThrAspGluThrMetVal 36
Qy      88 GTCTCGGCATCCGGCTATGAGAAAAGCTGACTAAGCGCCCGCAGTGTCTTGTGATT 147
Db      37 ValThrAlaAlaGlyTyrAlaGlnValIleGlnAlaProAlaSerIleSerValIle 56
Qy      148 AGCCAGGAGGAATTCAGTCCAGCCAGTACACAGTCTGGCGGAGGCTCTGAGATCAGTA 207
Db      57 SerArgGluAspLeuGluSerArgTyrArgAspValThrAspAlaLeuLysSerVal 76
Qy      208 GAGGTTGTGATGTTGAAAGTGTACGGGTAAACCGGAGGGCTGGAATACAGATCCGA 267
Db      77 ProGlyValThrValThrGlyGlyAspThr-----AspIleSerIleArg 93
Qy      268 GGAATGCCAGCCAGTACACGCTGATATGATGTTCTCTCAGGCGCGGAAGCAGT 327
Db      94 GlyMetGlySerAsnTyrThrLeuIleValAspGlyLysArgGln---ThrSerArg 112
Qy      328 GACGTGACTCCCAACGGTTTTCT---GCCATGATACCGGGTTCATGCCCTCTCGCC 384

```

Db	427	GluValThrProAspTrpGlyGlnValSerGlyGlyGly-----AsnIleTyrGlyAsn	444
Qy	1402	CCCGACCTGAAGCCGGAAGAGAGCGTCAGTTATCAGGCTGGGTGTATTACGATAACCCC	1461
Db	445	ProAspLeuGlnProGluThrSerIleAsnLysGluLeuSerLeuMetTyrSerThrGly	464
Qy	1462	CCCGGTCTGAATGCCAATGTCACAGGTTTATAGCTGACTTCTCCAAAGAGATTGTCTCT	1521
Db	465	SerGlyLeuAlaAlaSerLeuThrAlaPheHisAsnAspPheLysAspLysIleThrArg	484
Qy	1522	TATTCCATAAATGATAAC-----ACCAATAGC	1548
Db	485	ValAlaCysProAlaAsnIleCysThrAlaGlyProAsnGlnTrpGlyAlaThrProThr	504
Qy	1549	TAT---GTAAACACGGAAGGCCGGTTCACAGGTGTGGAATTTGCGGCACATTGCCG	1605
Db	505	TyrArgValAsnIleAspGluAlaGluThrTyrGlyAlaGluAlaThrLeuSerLeuPro	524
Qy	1606	CTGTGGTCAGAGATGTCACGCTGTCTACTGAATTACACCTGGAGCCCGAAAGTGAACACCT	1665
Db	525	Ile---ThrGluSerValGluLeuSerSerTyrThrTyrThrHisSerGluGlnLys	543
Qy	1666	GATGGTGATACAAAGTGGCCGCTCAGTTATACCCCTCAACACATGGTGATGCGAAA	1725
Db	544	SerGlyAsnPheAlaGlyArgProLeuLeuGlnLeuProLysHisLeuPheAsnAlaAsn	563
Qy	1726	CTGAACCTGGCAGATCACCGAAGAGGTGGCATCATGGCTGGGTCCCGCTTATCCGCGAAA	1785
Db	564	LeuSerTrpGlnThrThrAspArgLeuAsnSerTyrAlaAsnLeuAsnTyrArgGlyLys	583
Qy	1786	ACACCAGTTTACCACCAATATTTCGTCTACTGAGCGCTGTACAGAAGAAGTGATGAT	1845
Db	584	-----GluMetGlnProGluGlyGlyAla	591
Qy	1846	GAGAAAGGAGAAATACCTG---AAAGCTGGACGGTGTGGATCGACGTCTGCTGGGAAG	1902
Db	592	SerAsnAspPheIleAlaProSerTyrThrPheIleAspThrGlyValThrTyrAla	611
Qy	1903	ATCAGCATCCCTGACGTGAATGCTGGCGGTGAATAACCTGTCTCAACAGGATTACAGT	1962
Db	612	LeuThrAspThrAlaThrIleLysAlaAlaValTyrAsnLeuPheAspGlnGluValAsn	631
Qy	1963	GAGTGAGCCTGTACAGTGCCGGTAAGATPACGCTGTATGCCGGTGATTACTTCCAGCG	2022
Db	632	-----TyrAla---GluTyr	635
Qy	2023	GGATCATCAACACAGGATATGTGATACCTGAGCGAAATTTACTGGATGTGCGTGAACAT	2082
Db	636	-----GlyTyrValGluAspGlyArgArgTyrTrpLeuGlyLeuAspIle	650
Qy	2083	CAGTTC	2088
Db	651	AlaPhe	652

```

US-09-543-681A-7936
; Sequence 7936, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS
; FILE REFERENCE: 2709,1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 7936
; LENGTH: 677
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-7936

```

Alignment Scores:		6.41e-84	Length:	677
Pred. No.:	Score:	1007.50	Matches:	252
Percent Similarity:	Conservative:	50.74%	Mismatches:	123
Best Local Similarity:	Indels:	34.10%	Gaps:	23
Query Match:		27.00%		
DB:		4		
US-10-625-972-4 (1-2091) x US-09-543-681A-7936 (1-677)				
QY	16	CTGGCTCCGCTAGTCACTCTCTCGGATTTTCAGCCAGCAGCATAGCTGCTGCAGAG	75	
DB	18	IleAlaLeuGlyValValAlaLeuAlaLeuAlaLeuSerPheValMetThrAlaSerAlaGlu	37	
QY	76	GAT-----GTGATGATTGCTCGGCATCCGGCTATCAGAAAAAGCTGACTACGCA	126	
DB	38	AspProGluArgLeuLeuValThrThrAlaSerGlyPheLysGlnThrValGluAspAla	57	
QY	127	GCAGCCAGTCTTCTGATTAGCAGGAGGAATTGCAGTCCAGCCAGTACACGATCTG	186	
DB	58	ProLaserValSerValThrArgGluGlnLeuGluThrLysSerTyrArgAspVal	77	
QY	187	GCAGAGCTCTGAGATCAGTAGAGGTGTGGATGTTGAAAGTGGTACGGTAAACCGGA	246	
DB	78	ThrAspAlaLeuLysAspValProGlyValLeuVal-----ThrGlyGlySer	94	
QY	247	GGGTGGAATCAGATCCAGGAATCCAGCCAGTTCACAGCTGATACCTGATGATTGATGT	306	
DB	95	SerSerAspIleSerIleArgGlyMetAspProLysTyrThrMetIleLeuValAspGly	114	
QY	307	GTTCGTGAGGCGGAGCAGTACGTGACTCCCAACCGT---TTTCTGCCATGAATACC	363	
DB	115	LysArgVal---AlaSerArgGluThrArgProAsnSerAspAsnSerGlyIleGluGln	133	
QY	364	GGGTTCATGCCCTCTGCGCCGATTCAGGATTCAGCTTATCAGGGGCGGATGCTCC	423	
DB	134	GlyTrpLeuProLeuProAlaIleGluArgIleGluValArgGlyProMetSer	153	
QY	424	ACACTGTATGCTCTGATCGATCGGCGGTGTGGATATCATTCACAGAAAGATGCA	483	
DB	154	SerLeuTyrGlySerAspAlaMetGlyGlyValIleAsnIleThrArgLysAlaGln	173	
QY	484	GACAAATGGCTCTTCGTCGTCATGCGGGCTGAATCTGCAGGAACACAAATGGGT	543	
DB	174	LysGluTrpAsnPheSerLeuArgGlyAspThrThrLeuThrGluArgLysAsnGluGly	193	
QY	544	AACAGCCAGCTTAATTTCTGGAGCAGTGCCTCTGTGGATGATTCGTGAGCTG	603	
DB	194	AsnThrGlyGlySerPheTyrAlaAlaGlyProLeuIleAspAsnValLeuGlyLeu	213	
QY	604	CAGGTACCGGTAGCACACACAGCGTTCAGGTTTCATCGGTACATCAGTACGATACA	663	
DB	214	LysLeuGlnGlyGlnTyrSerHisArg-----Gly	223	
QY	664	CGAGCCAGCTATTCCTTATCCACGAGTCACAGAAATTAATCTTGTGTCAGCTT	723	
DB	224	GluAspArgArgValAspGlyTyrAsnArgGlnIleThrAlaSerGlyGlyLeu	243	
QY	724	GACTCGAAGCGTCCGACGAGATGCTCTGTTTGTATGATGATACACCCGCGAGCGT	783	
DB	244	SerTrpThrProAspGlyLysAsnThrValGluPheGluPheLysAspAsnGlnHis	263	
QY	784	TATGATAACCGGATGGGCAACTGGGAGTCTGACGGGGGATATGACCGACCTGGCG	843	
DB	264	ArgAspSerArg-----ValGlyTyrSerLysSerAlaGlu	275	
QY	844	-----TATGCGGAAACAAATTTCA	864	
DB	276	LeuValLysGlyLysAlaProGlnSerSerPheThrAspTyrGluLeuThrHisTyrAla	295	
QY	865	GCTGGCTATGATCATCTTACCTTCGGAACATCGAATCTGTAATCGAATCGAACGAG	924	

DB	296	LeuThrHisAspGlyIleTyrAspPheGlyThrMetAsnThrTyrValGlnArgAspGlu	315	
QY	925	ACAGAAATTAAGCTCGTGCAGCTTGTACGCAGCTGTACTGAAGCGCGCAAAATGGGGCTT	984	
DB	316	SerArgAsnProSerArg-----LysMetAsp-----	324	
QY	985	CGCGTTCAGCCGCGGAGCTTAAGAACTCGAACTTATCTCGAATCTATTAATCTGCTTACC	1044	
DB	325	-----TyrAspAspThrThrValArgAsnGlnThrValPhe---	336	
QY	1045	CCTCTGGAGAAATCTCATCTCGTTACGGTGGGCGAGTTTCAGAGCTCGCTCCATGAAA	1104	
DB	337	---MetPheGlyAspHisMetLeuSerValGlyGlyGlnTyrArgTyrGluGluLeuLys	355	
QY	1105	GAC---GGAGTGTCTCTCCAGCAGCAGGTGAACTTCCCGCAGAAAAGCTGTCGCTGA	1161	
DB	356	AspGluGlyAsnLysLeuLysGlyAsnLysLeuAspArgTyr---SerTrpAlaLeu	374	
QY	1162	TTTGTGAGGATGAGTGCATCTCAGCATGCATCTCGCTGCGCTGCATATCTGCTCGGATGTG	1221	
DB	375	PheAlaGluAspGluTrpThrLeuThrAsnAspPheAlaLeuThrAlaGlyLeuArgMet	394	
QY	1222	GAACATCATGACCAATTCGGGGGACATTCAGTCCGCTGCATATCTGCTCGGATGTG	1281	
DB	395	AspLysAspGluAsnPheGlyThrHisTrpThrProArgValTyrGlyValTrpHisLeu	414	
QY	1282	GCAGATGCTGCACCTGAAAGCGGTGTGACCGGATATAGCCAGATATAGCCAGATGGG	1341	
DB	415	AlaAspGluTrpThrLeuLysGlyValSerThrGlyTyrArgSerProAspLeuArg	434	
QY	1342	CAGCTACATAAAGGATAGTGTGTCTCGGGCAGGAGGAAAACAAAT-----CTACTT	1395	
DB	435	GlnAlaThrAlaThrTrpGlyGlnGlyThrGlyGlyArgTyrAspAlaValIleTyr	454	
QY	1396	GGTAACCCCGACTGAAGCCGAGAGCGCTCATGATTAGAGCTGGGTGATTACGAT	1455	
DB	455	GlyAsnProAspLeuLysProGluLysSerValThrGluGluIleGlyIleTrpAsp	474	
QY	1456	AACCCCGCGCTGAATGCCAATCTCAGAGTTTATGACTGCTTCCCAACAGATT	1515	
DB	475	AsnArgAspAsnLeuThrAlaSerLeuThrIleTyrAsnThrAspPheLysAspLysIle	494	
QY	1516	GTCTCTTAT-----TCCATAAATGATAACACCAATAGCTATGTA-----	1554	
DB	495	ThrGluValArgArgCysAspSerArgSerAspLysAlaProSerMetLeuAspCysThr	514	
QY	1555	-----AACAGCGGA-----	1563	
DB	515	LeuAlaAspGlyValGlyAsnSerGlyAspProTyrAspPheValSerAspArgThrAsn	534	
QY	1564	-----AAGCCCGGTTCACGCTGTGGAATTTCCGCGCACATTGCGCTGTGG-----	1611	
DB	535	ValAspLysAlaAsnMetArgGlyValGluValThrAlaAsn-----TrpIleIle	551	
QY	1612	TCAGAGATGTACGCTGTCATCTGAATTAACCTGACCCGAGAGTGAACACGATGATGT	1671	
DB	552	SerProGluTrpAsnLeuAlaAlaAsnTyrThrPheThrAspThrGluGlnLysSerGly	571	
QY	1672	GATAACAAAGGTCCGCGCTGATTAACCTCGAACACATGCTGAATCGCAAACTGAAC	1731	
DB	572	AspPheLysGlyLysProLeuAsnLysGlnProArgHisMetAlaAsnAlaThrLeuAsn	591	
QY	1732	TGGCAGCATCACCAAGAGGTGGCATCATGCTGGTCCCTGTTATCCGCGGAAAAACACCA	1791	
DB	592	TrpGluThrThrProGluMetGluThrTrpAlaArgIleAsnPheArgGlyLys-----	609	
QY	1792	CGTTTCCAGCAATATTCTGCTACCTGAGCGGTGTACAGAAAGTGTATGATGAGAAA	1851	
DB	610	-----ThrSerAspTyrLeuSerArgThrSerMetSerLysSerThr-----	623	
QY	1852	GGAGAAATACCTGAAAGCTCGACGGTGTGGATGTCAGGCTGTGCTGCGGAGATGACCGAT	1911	
DB	624	-----ProSerTyrAlaPheValAspValGlyThrSerTyrSerLeuThrLys	639	

Db 30 GlyThrMetValValThrAlaSerSerValGluGlnAsnLeuLysAspAlaProAla 49
 Qy 133 AGTGTTCTGTGATTAGCCAGGAAATTCAGTCCAGCCAGTACCAGTCTCGCGAG 192
 Db 50 SerIleSerValIleThrGlnLysLeuGlnArgLysProValGlnAsnLeuLysAsp 69
 Qy 193 GCTCTGAGATCAGTAGAGGTGTGGATGTTGAAGTGTGACGGTAAACCGGAGGCTG 252
 Db 70 ValLeuLysGluValProGlyValGlnLeuThrAsnGluLysAspAsnArgLysGlyVal 89
 Qy 253 GAAATCAGATCCAGAGGATCCAGCCAGTATACCGTACGTTACGTTGATGTTGCT 312
 Db 90 *****SerIleArgGlyLeuAspSerSerTyrThrLeuIleValAspGlyLysArg 109
 Qy 313 CAGGCGGAAGCAGTACGCTACCTCCCAACGGTCTTTCTGCAATCAATACCGGCTCATG 372
 Db 110 ValAsnSerArgAsnAlaValPheArgHisAsnAspPheAspLeuAsnTrpIle***** 129
 Qy 373 CCCCTCTGCGCCATTCAGCGTATTCAGGTTATCAGGGCCGCGATGTCACACTGAT 432
 Db 130 ***ProValAspSerIleGluArgIleGluValValArgGlyProMetSerSerLeuTyr 149
 Qy 433 GCGTCTGATCGGATGGCGGTGTGGTGAATATCATATACAGAAAGAAATGCAGACAAATGG 492
 Db 150 GlySerAspAlaLeuGlyGlyValValAsnIleIleThrLysLysIleGlyLysTrp 169
 Qy 493 CTCCTCTCCGCAATGACGGCTGAATCTGCAGGAAGCAACAAATGGGTAAACAGCAGC 552
 Db 170 SerGlyThrValThrValAspThrThrIleGlnGluHisArgAspArgGlyAspThrTyr 189
 Qy 553 CAGTTAATTTCTGGAGCAGTGGTCCCTCTGATGATGATCTGTCAGCGTCCAGTACGC 612
 Db 190 AsnGlyGlnPhePheThrSerGlyProLeuIleAspGlyValLeuGlyMetLysAlaTyr 209
 Qy 613 GGTAGCACACACACGCT-*****CAGGTTCTATCGGTCACTACTACG 657
 Db 210 GlySerLeuAlaLysArgGlyLysAspAspProGlnAsnSerThrThrAspThrGly 229
 Qy 658 GATACAGCAGGACCGGTATTCCTTATCCACGGAGTCCAGAAATTAATCTGTGTGCA 717
 Db 230 GluThr-----Pro 232
 Qy 718 GCTCTGTGATGAAGCGTCGAGCAG-----GATGTGCTCTGTTGATATGAT 768
 Db 233 ArgIleGluGlyPheSerSerArgAspGlyAsnValGluPheAlaTrp----- 248
 Qy 769 ACCACCCGCGCGCTTATGATAACCGGATGGCACTGGGAGTCTGACGGGGGATAT 828
 Db 249 ThrProAsnGlnAsnHisAspPheThrAlaGlyTyr-----GlyPhe 262
 Qy 829 GACCGG----- 834
 Db 263 AspArgGlnAspArgAspSerAspSerLeuAspLysAsn***** 282
 Qy 835 ---ACCTGCGCTATGACGGAACAAATTTTCAGTGGCTATGATCATCTTTTCACTTC 891
 Db 283 *****ArgLeuGluArgGlnAsnTyrSerValSerHisAsnGlyArgTrpAspTyr 302
 Qy 892 GGAACATGGAATCTCTGACTGACCTGACGAGCAGAGAAATTAAGTCTGCTGAGCTGTA 951
 Db 303 GlyThrSerGluLeuLysTyrTyrGlyGluLysValGluAsnLysAsn***** 322
 Qy 952 CGCAGTGTACTGAAGCGGACAAATGGGGGCTTTCGCGGTGACGCGCGGAGCTTAAGGAA 1011
 Db 323 *****ProGlyAsnSerSerProIleThrSerGlu 342
 Qy 1012 TCGAACCTTATCTGAATTCATCTGCTTACCTCTGGAGAAATCTCATCTGTTACG 1071
 Db 343 SerAsnThrValAspGlyLysTyrThrLeuProLeuThrAlaIleAsnGlnPheLeuThr 362
 Qy 1072 GTGGGGGCGAGTTTCAGAGCTCGCTCCATGAAGACGAGTGTCTCTTCCGCCACACAGGT 1131

Db 363 ValGlyGlyGluTrpArgHisAspLysLeuSerAspAlaValAsnLeu-----ThrGly 380
 Qy 1132 GAACTTTTCGCGCAGAAAAAGC-----TGCTCGTATTCTCTGAGGATGAGTGGCAT 1182
 Db 381 GlyThrSerSerLysThrSerAlaSerGlnTyrAlaLeuPheValGluAspGluTrpArg 400
 Qy 1183 CTCACGATGACCTTCGCTGATCGGCGACCGCTATGAACATCATCAGCAATTCGGG 1242
 Db 401 IlePheGluProLeuAlaLeuThrThrGlyValArgMetAspHisGluThrTyrGly 420
 Qy 1243 GGAACCTTCAGTCCGCGTGCATATCTGCTCGGATGTGGCAGATCCCTGAGCGCTGAAA 1302
 Db 421 GluHisTrpSerProArgAlaTyrLeuValTyrAsnAlaThrAspThrValThrValLys 440
 Qy 1303 GCGGTGTGACACCGGATATAAGGACCCAGATGGGCGACGCTACATAAAGGATTAT 1362
 Db 441 GlyGlyTrpAlaThrAlaPheLysAlaProSerLeuLeuGlnLeuSerProAspTrpArg 460
 Qy 1363 GGTGTCTCGCGCAGGAAAAAACAATCTACTTGTGAACCCGACCTGAAGCCGGAAGAG 1422
 Db 461 SerAsnSerCysArgGlyAlaCysLysIleValGlySerProAspLeuLysProGluThr 480
 Qy 1423 AGCGTCAGTTATGAGCTCGGTGTATTACGATAACCC-----GCCGTCTG 1470
 Db 481 SerGluSerTrpGluLeuGlyLeuTyrTyrMetGlyGluGlyTrpLeuGluGlyVal 500
 Qy 1471 AATGCCAAATGTCACAGGTTTATGACTGCTCTCCAAACAGATGCTCTTATTCATTA 1530
 Db 501 GluSerSerValThrValPheArgAsnAspValLysAspArgIle---SerIleSerArg 519
 Qy 1531 AATGATAACACCAAT----- 1545
 Db 520 ThrSerAspValAsnAlaAlaProGlyTyrGlnAsnPheValGlyPheGluThrGlyAla 539
 Qy 1546 -----ACCTATGTAACAGCGGAAGCGCGGTGTCAGCT 1581
 Db 540 AsnGlyArgArgIleProValPheSerTyrTyrAsnValAsnLysAlaArgAsnGlnGly 559
 Qy 1582 GTGCAATTTCCGCGCATTGCGCTGTGTGAGAGATGTCAGCTGTCACCTGCTGAATTAC 1641
 Db 560 ValGluThrGluLeuLysIlePro---PheAsnAspGluTrpLysLeuSerIleAsnTyr 578
 Qy 1642 ACTGACCCGAGAGTGAACACAGTGAT-----GGTGATAACAAAGTGGCGCTG 1692
 Db 579 ThrTyr-----AsnAspGlyArgAspValSerAsnGlyGluAsnLys-----ProLeu 594
 Qy 1693 AGTTATACCCCTCAACACATGGTGAATGCGAACTGAACCTGCGCAGATCACCGAAGAGTG 1752
 Db 595 SerAspLeuProPheHisLeuAla-----LeuGluAspTrp 606
 Qy 1753 GCATCATGTGGTGGTCCCGTTATCGCGGAAAAACACCGTTTCCACCGAATTTATCG 1812
 Db 607 SerPheTyrValSerGlyHisTyrThrGly----- 616
 Qy 1813 TCACTGAGCGCTACAGAAAGTGTATGATGAGAAAGGAGATACCTGAAGCGCTG 1872
 Db 617 -----GlnLysArgAlaAspSerAlaThrAlaLysThrProGlyGlyTyr 631
 Qy 1873 ACGGTGTGTGATCAGCTCTGCTGGAAGATGACCGATGCCCTGACGCTGTAATGTGCG 1932
 Db 632 ThrIleTrpAsnThrGlyAlaAlaTrpGlnValThrLysAspValLysLeuArgAlaGly 651
 Qy 1933 GTGAATACCTGTCTCAACAGGATTCAGTACGCTGAGCTGTACAGTCCGCT----- 1986
 Db 652 ValLeuAsnLeuGlyAspLys***** 671
 Qy 1987 -----AAGAGTACGCTGATGCGGTGATTACTTCCAGACGGGATCA 2028
 Db 672 AsnGlyThrLeuAspTrpLysProAspLeuSerArgAspLys----- 686
 Qy 2029 TCACCAACAGGATATGTATACCTGACGGAATTTACTGGATGCTGCTGCAATCTCATGTC 2088
 Db 687 -----SerTyrAsnGluAspGlyArgTyrPheMetAlaValAspTyrArgPhe 703

RESULT 13

US-08-765-081-7
 : Sequence 7, Application US/08765081
 : Patent No. 5798260
 : GENERAL INFORMATION:

APPLICANT: Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
 TITLE OF INVENTION: Escherichia Coli 0157:H7 Epithelial Adhesin
 NUMBER OF SEQUENCES: 12
 CORRESPONDENCE ADDRESS:

ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
 STREET: 2800 Pacific First Center, 1420 Fifth Avenue
 City: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98101-2347

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette-3.5 inch, 1.44Mb storage

COMPUTER: IBM PC compatible/Pentium

OPERATING SYSTEM: MS-Windows 3.1

SOFTWARE: Word for Windows-6.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/765,081

FILING DATE: March 26, 1997

PRIOR APPLICATION DATA: PCT/US95/06994

FILING DATE: June 7, 1995

APPLICATION NUMBER: US 08/265,714

FILING DATE: June 24, 1994

ATTORNEY/AGENT INFORMATION:

NAME: Shelton, Dennis K.

REGISTRATION NUMBER: 26,997

REFERENCE/DOCKET NUMBER: CHOR-1-10286

TELECOMMUNICATION INFORMATION:

TELEPHONE: 1-206-682-8100; 1-206-224-0718 (direct)

TELEFAX: 1-206-224-0779

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 663 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

DESCRIPTION: E. coli CtxA protein amino acid sequence

HYPOTHETICAL: NO

ORIGINAL SOURCE:

ORGANISM: Escherichia Coli

US-08-765-081-7

Alignment Scores:

Pred. No.:	6,93e-65	Length:	663
Score:	800.50	Matches:	231
Percent Similarity:	47.18%	Conservative:	120
Best Local Similarity:	31.05%	Mismatches:	242
Query Match:	21.45%	Indels:	151
DB:	1	Gaps:	25

US-10-625-972-4 (1-2091) x US-08-765-081-7 (1-663)

QY	40	CTCGGATTTCAGCCAGCAGCAGTAGCTGCTGCA-----	72
Db	10	ValGlyLeuCysLeuSerAlaIleSerCysAlaTrpProValLeuAlaValAspAsp 29	
QY	73	GAGGATGTGATGATTGTCTCGGCATCCGGCTATGAGAAAAAGCTGACTAACGCGGCC 132	
Db	30	GlyGluThrMetValValThrAlaSerValGluGlnAsnLeuLysAspAlaProAla 49	
QY	133	AGTGTTCTGTGATAGCAGCAGGAAATTCAGTCCAGCCAGTACCATCTGCGGAG 192	
Db	50	SerIleSerValIleThrGlnGluAspLeuGlnArgLysProValGlnAsnLeuLysAsp 69	
QY	193	GCTCTGAGTAGTAGAGGTGTGGATGTT---GAAAGTGGTACGGGTAAACCGCGAGG 249	
Db	70	ValLeuLysGluValProGlyValGlnLeuThrAsnGluGlyAspAsnArgLysGly--- 88	

QY	250	CTGGAATACGATCCGAGGAATCCAGCAGTTACAGCTGATGATGATGATGTT 309
Db	89	-----ValSerIleArgGlyLeuAspSerSerThrLeuIleLeuValAspGlyLys 106
QY	310	CGTCAGGCGGGAAGCAGTACGTGACTCCCAAGGGTTTCTGCCATGATACCGGGTTC 369
Db	107	ArgValAsnSerArgAsnAlaValPheArgHisAsnAspPheAspLeuAsn----- 123
QY	370	ATGCCCTCTCTGCCGCAATTGAGCGTATTGAGTTATCAGGGGCGCGATGTCACACTG 429
Db	124	TrpIleProValAspSerIleGluArgIleGluValValArgGlyProMetSerLeu 143
QY	430	TATGGCTCTCATCGATGGCGGTGGTGAATATCATACGAAAGAAATGACAGCAA 489
Db	144	TyrGlySerAspAlaLeuGlyValValAsnIleThrLysLysIleGlyGlnLys 163
QY	490	TGGCTCTCTCCGTCATGCAGGCTGAATCTCCAGGAAAGCAAAATGGGTACAGC 549
Db	164	TrpSerGlyThrValThrValAspThrThrIleGlnGluHisArgAspArgGlyAspThr 183
QY	550	AGCCAGTTTAATTTCTGAGCAGTGTCCCTTGTGGATGATCTCTCAGCCCTCAGGTA 609
Db	184	TyrAsnGlyGlnPheThrSerGlyProLeuIleAspGlyValLeuGlyMetLysAla 203
QY	610	CGCGGTAGCACACACAGCGT-----CAGGGTTTCATCGGTCACTCAGCTG 654
Db	204	TyrGlySerLeuAlaLysArgGluLysAspAspProGlnAsnSerThrThrAspThr 223
QY	655	AGCGATACAGCAGCGCGGTATTCTTATCCACGAGTACAGAAATTATAATCTTGT 714
Db	224	GlyGluThr-----
QY	715	GCAGTCTTCAGTGGAGGCTCGAGCAG-----GATGTGCTCTGGTTGATATG 765
Db	227	ProArgIleGluGlyPheSerSerArgAspGlyAsnValGluPheAlaTrp----- 243
QY	766	GATACACACCGCGCGGTATTGATAACCGGGATGGGCAACTGGGGAGTCTGACGGGGA 825
Db	244	--ThrProAsnGlnAsnHisAspPheThrAlaGlyTyr-----Gly 256
QY	826	TATGACCG-----ACCTGGCTATGAGGAAACAA 858
Db	257	PheAspArgGlnAspArgAspSerAspSerLeuAspLysAsnArgLeuGluArgLys 276
QY	859	ATTTGAGTGGCTATGATCATCTTTACCTTCGGAACATGGAATCGTATCTGAACGG 918
Db	277	TyrSerValSerHisAsnGlyArgTrpAspTyrGlyThrSerGluLeuLysTyrTyrGly 296
QY	919	AACGAGACAGAAATAAAGTTCGTGAGCTTGTCAGCAGTGTACTGAAGCGCAGCAATGG 978
Db	297	GluLysValGluAsnLysAsn-----
QY	979	GGCTTCCCGCTCAGCGCGGGAGCTT---AAGGAATCGAATCTTATCTGATTCATTA 1035
Db	304	-----ProGlyAsnSerSerProIleThrSerGluSerAsnThrValAspGlyLysTyr 321
QY	1036	CTGCTTACCCCTCTGGGAGAACTCTCATCTGGTTACGTTGGGGGCGAGTTTCAGAGCTCG 1095
Db	322	ThrLeuProLeuThrAlaIleAsnGlnPheLeuThrValGlyGlyGluTrpArgHisAsp 341
QY	1096	TCCATGAAGACGGAGTTGTCCTTCCAGCAGCAGGTGAACCTTCGGGAGAAAGC--- 1152
Db	342	LysLeuSerAspAlaValAsnLeu-----ThrGlyGlyThrSerSerLysThrSerAla 359
QY	1153	-----TGTGGTATTTCTGAGTAGTGAGTCTCACGATGCTCACGATGCTGCTGACT 1206
Db	360	SerGlnTyrAlaLeuPheValGluAspGluTrpArgIlePheGluProLeuAlaLeuThr 379
QY	1207	GCGGCGAGCGGTATGAACATATGAGCAATTCGGGGGACACTTCAGTCCGGCTCATAT 1266
Db	380	ThrGlyValArgMetAspAspHisGluThrTyrGlyGluHisTrpSerProArgAlaTyr 399


```

QY 1267 CTGGCTGGGATGGGAGATGCTGGACGCTGAAGCGGTGTGACACGGGATATAAG 1326
Db 400 LeuValTyrAsnAlaThrAspThrValThrValLysGlyGlyTrpAlaThrAlaPheLys 419
QY 1327 GCACCCAGAAATGGGCGACCTACATAAAGGATATTAGTGTGTGTCGGGCGAGGAAAAACA 1386
Db 420 AlaProSerLeuLeuGlnLeuSerProAspTrpThrSerAsnSerCysArgGlyAlaCys 439
QY 1387 AATCTACTGTGTAACCCGACCTGAAGCCGGAAGAGCGTCAGTTATGAGCGTGGGTG 1446
Db 440 LysIleValGlySerProAspLeuLysProGluThrSerGluSerTrpGluLeuGlyLeu 459
QY 1447 TATTACGATAACCC-----GCCGGTCTGAATGCCAATGTCACAGGTTTATG 1494
Db 460 TyrTyrMetGlyGluGluGlyTrpLeuGluGlyValGluSerSerValThrValPheArg 479
QY 1495 ACTGACTTCTCCACAAAGATTGCTCTTATTTCCTAATAATGATAACCAAT----- 1545
Db 480 AsnAspValLysAspArgile---SerIleSerArgThrSerAspValAsnAlaPro 498
QY 1545 ----- 1545
Db 499 GlyTyrGlnAsnPheValGlyPheGluThrGlyAlaAsnGlyArgGlyProValPhe 518
QY 1546 AGCTATGTAACAGCGGAAGCCCGGTGTCACGCTGTGGAATTTGCCGACATTCGCG 1605
Db 519 SerTyrTyrAsnValAsnLysAlaArgAsnGlnGlyValGluThrGluLeuLysIlePro 538
QY 1606 CTGTGCTCAGAGGATGTCACCTGCTCACTGTAATACACCTGACCCGAGTGAACACGT 1665
Db 539 ---PheAsnAspGluTrpLysLeuSerIleAsnTyrThrTyr-----AsnAspGlyArg 555
QY 1666 GAT-----GGTGATAACAAAGTGGCGCGTGTAGTTATACCCCTGAACACATGGTG 1716
Db 556 AspValSerAsnGlyGluAsnLys-----ProLeuSerAspLeuProPheHisLeuAla 573
QY 1717 AATGCGAACTGAACCTGCGCAGATCACCGAAGAGTGTCATCGCTGGTGGTCCGCTAT 1776
Db 574 -----LeuGluAspTrpSerPheTyrValSerGlyHisTyr 585
QY 1777 CGCGGAAACACACCGATTTCACCCAGAAATTTCTGTCACCTGAGCGGTGTACAGAA 1836
Db 586 ThrGly-----GlnLysArg 590
QY 1837 GTGTATGATGAGAAGAGAGATACCTGAAGCCCTGGACGGTGGTGGATGACAGTCTGCG 1896
Db 591 AlaAspSerAlaThrAlaLysThrProGlyGlyTyrThrIleTrpAsnThrGlyAlaAla 610
QY 1897 TGGAGATGACGGATGCCCTGACGCTGAATGCTGGGTGAATAACCTGCTCAACAGGAT 1956
Db 611 TrpGlnValThrLysAspValLysLeuArgAlaGlyValLeuAsnLeuGlyAspLysThr 630
QY 1957 TACAGTACGTGAGCCTGTACAGTCCCGGTAAAGATACGCTGTATGCGGTGATTACTTC 2016
Db 631 AlaAsnGlyThrLeuAspTrp-----LysProAspLeuSerArgAspLysTyr--- 646
QY 2017 CAGCGGATCATCAACACAGGATATGTATACCTGAGCGAAATTAAGTATGATGCTGCTG 2076
Db 647 -----SerTyrAsnGluAspGlyArgArgTyrPheMetAlaVal 659
QY 2077 AACTATCAGTTC 2088
Db 660 AspTyrArgPhe 663

```

RESULT 14

```

US-09-098-082-7
; Sequence 7, Application US/09098082
; Patent No. 6040421
; GENERAL INFORMATION:
; APPLICANT: Barr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
; TITLE OF INVENTION: Escherichia Coli 0157:H7 Epithelial Adhesin
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:

```

```

; ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
; STREET: 2800 Pacific First Center, 1420 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette-3.5 inch, 1.44Mb storage
; COMPUTER: IBM PC compatible/Pentium II
; OPERATING SYSTEM: MS-Windows 95
; SOFTWARE: Word for Windows-6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/098,082
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/765,081
; FILING DATE: March 26, 1997
; APPLICATION NUMBER: PCT/US95/06994
; FILING DATE: June 7, 1995
; APPLICATION NUMBER: US 08/265,714
; FILING DATE: June 24, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Sheiness, Diana K.
; REGISTRATION NUMBER: 35,356
; REFERENCE/DOCKET NUMBER: CHOR-1-12402
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 1-206-682-8100; 1-206-224-0735 (direct)
; TELEFAX: 1-206-224-0779
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 663 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; DESCRIPTION: E. coli CirA protein amino acid sequence
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Escherichia Coli
; US-09-098-082-7

```

```

Alignment Scores:
Pred. No.: 6 93e-65 Length: 663
Score: 800.50 Matches: 231
Percent Similarity: 47.18% Conservative: 120
Best Local Similarity: 31.05% Mismatches: 242
Query Match: 21.45% Indels: 151
DB: 3 Gaps: 25

```

```

US-10-625-972-4 (1-2091) x US-09-098-082-7 (1-663)
QY 40 CTCGGATTTTCACCCAGCAGCATAGCTGCTGCA----- 72
Db 10 ValGlyLeuCysLeuSerAlaIleSerCysAlaTrpProValLeuAlaValAspAsp 29
QY 73 GAGAGTGTGATGATTCCTCGGCATCCGCTATGAGAAAAGCTGACTAACGCGCGCC 132
Db 30 GlyLThrMetValValThrAlaSerSerValGluAsnLeuLysAspAlaProAla 49
QY 133 AGTGTTCGTGTATGATCCAGGAGAAATTCAGTCCAGCCAGTACCAGATCTGCGCGAG 192
Db 50 SerIleSerValIleThrGlnGluAspLeuGlnArgLysProValGlnAsnLeuLysAsp 69
QY 193 GCTCTGAGTACGATGAGGCTGTGGATGT---GAAAGTGTACGGTAAACCGGAGG 249
Db 70 ValLeuLysGluValProGlyValGlnLeuThrAsnGluGlyAspAsnArgLysGly--- 88
QY 250 CTGGAATCAGCATCCGAGGAATGCCAGCAGTACACCGTACTGATGATGATGTT 309
Db 89 -----ValSerIleArgGlyLeuAspSerTyrThrLeuIleLeuValAspGlyLys 106
QY 310 CGTCAGGCGGAGAGCAGTACGTCATCCCAACGGTTTCTGCCATGATACCGGTTTC 369
Db 107 ArgValAsnSerArgAsnAlaValPheArgHisAsnAspPheAspLeuAsn----- 123

```

370 ATGCCCTCTGGCCGCCATTCAGCGCTATTGAGGTTATCAGGGGCGCGATGTCACACTG 429
 124 TptleProValaspSerlleGluArglleGluValValArgGlyProMetSerSerLeu 143
 430 TATGGCTCTGATCGCATGGCGGTGTGGTGAATATCATACCAGAAAGAATGCAGACAAA 489
 144 TyrGlySerAspAlaLeuGlyGlyValValAsnIleIleThrIlysIleGlyGlnLys 163
 490 TGGGTCTCTTCGGTCAATGCAGGCTCAATCTGCAGGAAGCAACAATGGGTGAACAGC 549
 164 TtpSerGlyThrValThrValAspThrThrIleGlnGluHisArgAspAsgGlyAspThr 183
 550 AGCCAGTTTAAATTCAGACAGTGTCCCTGTGTGATGATCTGTCTCAGCCTGCAGGTA 609
 184 TyrAsnGlyGlnPhePheThrSerGlyProIleuIleAspGlyValLeuGlyMetLysAla 203
 610 CGGGTAGCACACAAACAGCT-----CAGGGTTCATCGGTCAACATCACTG 654
 204 TyrGlySerLeuAlaLysArgGluLysAspAspProGlnAsnSerThrThrThrAspThr 223
 655 AGCGATACAGCAGGCACCGCTATTCCTTATCCACCGAGTCACAGAAATATAATCTTGGT 714
 224 GlyGluThr----- 226
 715 GCAGCTTCTTACTGGAAGCGTCGAGCAG-----GATGTGCTCTCGTTGATATG 765
 227 ProArgIleGluGlyPheSerSerArgAspGlyAsnValGluPheAlaItp----- 243
 766 GATACCAACCGGACCGCTTATGATACCGGATGGGCAACTGGGGAGCTGACGGGGGGA 825
 244 --ThrProAsnGlnAsnHisAspPheThrAlaGlyTyr-----Gly 256
 826 TATGACCG-----ACCTGCGCTATGAGCGMAACAAA 858
 257 PheAspArgGlnAspArgAspSerSerLeuAspLysAsnArgLeuGluArgGlnAsn 276
 859 ATTTCAGCTGCCTATGATCATCTTCACCTTCGGAACATGGAATCGTATCTGAACTGG 918
 277 TyrSerValSerHisAsnGlyArgTtpAspTyrGlyThrSerGluLeuLysTyrTyrGly 296
 919 AACGAGACAGAAATAAAGTCTGTGAGCTTGTACGCAGTCTACTGAAGCGGACAAATGG 978
 297 GluLysValGluAsnLysAsn----- 303
 979 GGGCTTCCCGCTCAGCCGCGGAGCTT--AAGAAATCGAACCTTATCTCTGAATTCATTA 1035
 304 -----ProGlyAsnSerSerProIleThrSerGluSerAsnThrValAspGlyLysTyr 321
 1036 CTGCTTACCCTCTGGGAGAATCTCATCTGGTACGCTGGGGCGGAGTTTCAGAGCTCG 1095
 322 ThrLeuProLeuThrAlaIleAsnGlnPheLeuThrValGlyGlyGluTtpAsgHisAsp 341
 1096 TCCATGAAGACGAGTTGCTTCCTGCCAGCACAGGTGAAACTTCCGGCGAGAAAGC--- 1152
 342 LysLeuSerAspAlaValAsnLeu-----ThrGlyGlyThrSerSerLysThrSerAla 359
 1153 -----TGGTCTGGTATTTCAGAGATGATGGGCATCTCCAGGATGCATTTGCGCTCACT 1206
 360 SerGlnTyrAlaLeuPheValGluAspGluTtpArgIlePheGluProLeuAlaLeuThr 379
 1207 GCGGGCAGCCGCTATGAACATCATAGCAATTCGGGGGACACTTCATCTCCGCTGCATAT 1266
 380 ThrGlyValArgMetAspAspHisGluThrTyrGlyGluHisTtpSerProAsgAlaTyr 399
 1267 CTGCTCTGGATGTGGCAGATGCTGTGACGCTGAAAGCGCGTGTGACACCGGATATAAG 1326
 400 LeuValTyrAsnAlaThrAspThrValThrValLysGlyGlyTrpAlaThrAlaPheLys 419
 1327 GCACCCAGAAATGGGGCAGCTACATAAAGGGATTAGTGGTGTGCCGGCAGGAAAAACA 1386
 420 AlaProSerLeuLeuGlnLeuSerProAspTtpThrSerAsnSerCysArgGlyValaCys 439

Qy	1387	AATCTACTTGTTAAACCCCGACCTGAACCCGGAGACAGCGCTCAGTTATGAGCGTGGGGTG	1446
Db	440	LysileValGlySerProAspLeuLysProGluThrSerGluSerTrpGluLeuGlyLeu	459
Qy	1447	TATTAGATAAACC	1494
Db	460	TyrTyrWecGlyGluGluGlyTrpLeuGluGlyValGluSerSerValThrValPheArg	479
Qy	1495	ACTGACTTCTCCAAACAGATTGTCTCTTATTCATAATAATGATAAACCAAT	1545
Db	480	AsnAspValLysAspArgile--SerIleSerArgThrSerAspValAsnAlaAlaPro	498
Qy	1545		1545
Db	499	GlyTyrGlnAsnPheValGlyPheGluThrGlyAlaAsnGlyArgArgIleProValPhe	518
Qy	1546	AGCTATGTAAACAGCGAAGGCCGGTTCACCGGTGTGGAATTTCCCGGCACATTCGCCG	1605
Db	519	SerTyrTrpAsnValAsnLysAlaArgAsnGlnGlyValGluThrGluLeuLysIlePro	538
Qy	1606	CTGTGTCAGAGGATGTCAGCTGTCTCACTGAATTACACTGGACCCGCAAGTGAACAACGT	1665
Db	539	---PheAsnAspGluTrpLysLeuSerIleAsnTyrThrTyr-----AsnAspGlyArg	555
Qy	1666	GAT-----CGTGATACAAAGGTGCCCGCTGAGTTATACCTCGTAACAACATGGTG	1716
Db	556	AspValSerAsnGlyGluAsnLys-----ProLeuSerAspLeuProPheHisLeuAla	573
Qy	1717	AATGCGAAACTGAACGGCAGATCACCGAAGAGGTGGCATCATGGTGGGTGCCCGCTTAT	1776
Db	574	-----LeuGluAspTrpSerPheTyrValSerGlyHisTyr	585
Qy	1777	CGCGGGAAACACACCGTTCCACCCAGATTATTCGTACTGAGCGCTGTACAGAGAA	1836
Db	586	ThrGly-----	590
Qy	1837	GTGTATGATGAGAAAGGAGAATACCTGAAAGCTGGACCGTGGTGGATGCAGGTCTGTGCG	1896
Db	591	AlaAspSerAlaThrAlaLysThrProGlyGlyTyrThrIleTrpAsnThrGlyAlaAla	610
Qy	1897	TGGAAGATGACGAGTCCCTGACGCTGAATGTGCGGTGAATAACTGTCTCAACAAGGAT	1956
Db	611	TrpGlnValThrLysAspValLysLeuArgAlaGlyValLeuAsnLeuGlyAspLysThr	630
Qy	1957	TACAGTGACGTGACCTGTACAGTGGCGGTAGAGTACGCTGTATCCCGGTGATTACTTC	2016
Db	631	AlaAsnGlyThrLeuAspTrp-----LysProAspLeuSerArgAspAspTyr---	646
Qy	2017	CAGACGGGATCATCAACACAGGATATGTGATACCTGACGGCAATTAACCTGGATGTGCGTG	2076
Db	647	-----SerTyrAsnGluAspGlyArgArgTyrPheMetAlaVal	659
Qy	2077	AACATACAGTTC	2088
Db	660	AspTyrArgPhe	663

```

RESULT 15
US-09-489-039A-8455
Sequence 8455, Application US/09489039A
Patent No. 6610836
GENERAL INFORMATION:
APPLICANT: Gary Breton et. al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 2709.2004001
CURRENT APPLICATION NUMBER: US/09/489,039A
CURRENT FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: US 60/117,747
PRIOR FILING DATE: 1993-01-29
NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 8455
LENGTH: 778
TYPE: PRT

```

APPLICANT: Gary Breton et. al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: October 13, 2004, 11:31:01 Search time 125 Seconds

(without alignments)

9452.909 Million cell updates/sec

Title: US-10-625-972-4

Perfect score: 3732

Sequence: 1 atcggaataaccactctggc.....cgctgaactatcagttctga 2091

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 3172214

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh
-Q=/cgn2_1/USPRO_pool/US10625972/runat_12102004_130105_26800/app_query.fasta_1.2247
-DB=A Geneseq 29Jan04 -QFMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bites -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCLALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10625972 -CGN 1_1_158 -runat_12102004_130105_26800 -NCPU=6 -ICPU=3
-NO_MMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

A Geneseq 29Jan04:*
1: Geneseqp1980s:*
2: Geneseqp1980s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3632	97.3	696	7 ADC00149	Adc00149 Enterohae
2	3628	97.2	696	2 AAR75366	Aar75366 Adhesin.
3	1029	27.6	660	6 ABM67028	Abm67028 Photorhab
4	972	26.0	682	4 ABBS2836	Abbs2836 Escherich
5	617.5	16.5	761	6 ADA34655	Ada34655 Acinetoba
6	604.5	16.2	725	4 AAY72914	Aay72914 E. coli 1
7	539.5	14.5	620	6 ABM68636	Abm68636 Photorhab
8	428.5	11.5	687	6 ABU20598	Abu20598 Protein e
9	399	10.7	684	6 ABU20991	Abu20991 Protein e
10	397.5	10.7	708	2 AAY34501	Aay34501 Porphorym

11	397.5	10.7	772	2 AAY34374	Aay34374 Porphorym
12	391	10.5	660	7 ADC01296	Adc01296 Enterohae
13	388	10.4	660	4 ABBS2669	Abbs2669 Escherich
14	383.5	10.3	643	6 ADA33859	Ada33859 Acinetoba
15	373.5	10.0	674	6 ABM67198	Abm67198 Photorhab
16	371	9.9	698	6 ABU49701	Abu49701 Protein e
17	365.5	9.8	663	4 ABBS2462	Abbs2462 Escherich
18	352.5	9.4	713	7 ADC01362	Adc01362 Enterohae
19	347.5	9.3	654	4 ABBS2577	Abbs2577 Escherich
20	340	9.1	704	6 ADA34250	Ada34250 Acinetoba
21	339	9.1	673	6 ABM67345	Abm67345 Photorhab
22	337.5	9.0	757	2 AAY34472	Aay34472 Porphorym
23	337.5	9.0	763	2 AAY34472	Aay34472 Porphorym
24	327.5	8.8	153	2 AAR60561	Aar60561 Vibrio ch
25	325.5	8.7	713	6 ABU49654	Abu49654 Protein e
26	321	8.6	775	6 ABU20485	Abu20485 Protein e
27	315	8.4	743	6 ADA33786	Ada33786 Acinetoba
28	314	8.4	428	6 ABP57973	Abp57973 Ferric-si
29	314	8.4	428	6 ABR38876	Abt38876 Surface e
30	313.5	8.4	151	2 AAR25587	Aar25587 Truncated
31	310.5	8.3	790	2 AAR95565	Aar95565 N. mening
32	304	8.1	791	2 AAR95566	Aar95566 N. mening
33	302.5	8.1	774	6 ABU28129	Abu28129 Protein e
34	302	8.1	729	6 ABU47292	Abu47292 Protein e
35	301	8.1	763	6 ABU20715	Abu20715 Protein e
36	300.5	8.1	774	4 AAU29336	Aau29336 Novel mar
37	300.5	8.1	774	6 ABU15308	Abu15308 Protein e
38	297.5	8.0	779	6 ABM68355	Abm68355 Photorhab
39	297	8.0	790	2 AAR95568	Aar95568 N. gonorr
40	296.5	7.9	722	3 AAB07697	Aab07697 A. Neisser
41	295	7.9	722	6 ABU27630	Abu27630 Protein e
42	294.5	7.9	766	6 ADA36943	Ada36943 Acinetoba
43	294	7.9	833	2 AAY34494	Aay34494 Porphorym
44	294	7.9	876	2 AAY34368	Aay34368 Porphorym
45	293.5	7.9	719	2 AAW01462	Aaw01462 NTHI HxuC

ALIGNMENTS

RESULT 1
ADC00149
ID ADC00149 standard; protein; 696 AA.
XX
AC ADC00149;
XX AC

DT 04-DEC-2003 (first entry)

DE Enterohaemorrhagic E. coli 0157:H7-specific protein SEQ ID NO: 194.

XX enterohaemorrhagic; anti-bacterial.

XX Escherichia coli; 0157:H7.

PN JP2002355074-A.

PD 10-DEC-2002.

XX 24-JAN-2002; 2002JP-00015959.

XX 24-JAN-2001; 2001JP-00112010.

PA (UYTS-) UNIV TSUKUBA.

XX WPI; 2003-451640/43.

XX Enterohaemorrhagic Escherichia coli 0157:H7-specific nucleic acid molecule and a polypeptide and its use, a polypeptide, a vector and a host cell.

XX Claim 3; SEQ ID NO 194; 2067bp; Japanese.

XX The invention relates to a novel enterohaemorrhagic Escherichia coli 0157:H7-specific nucleic acid molecule. A polynucleotide of the invention has anti-bacterial activity. The polypeptide can be used in detection

CC and/or treatment of O157:H7 infection. The nucleotide sequence of the
 CC genome of Enterohaemorrhagic E coli O157:H7 was determined. The present
 CC sequence represents an E. coli O157:H7-specific polypeptide of the
 CC invention.

XX
 SQ Sequence 696 AA;

Alignment Scores:

Pred. No.: 0 Length: 696
 Score: 3632.00 Matches: 695
 Percent Similarity: 99.86% Conservative: 0
 Best Local Similarity: 99.86% Mismatches: 1
 Query Match: 97.32% Indels: 0
 DB: 7 Gaps: 0

US-10-625-972-4 (1-2091) x ADC00149 (1-696)

QY 1 ATGCGAATACACACTCTGGCTTCCTAGTATCTCCCTCTCGATTTTCAGCCAGCAGC 60
 DB 1 MetArgIleThrLeuAlaSerValValIleProCysLeuGlyPheSerAlaSerSer 20
 QY 61 ATAGCTGCTGCAGAGATGTGATGATTGTCCTGGCATCCGGTATGAGAAAAGCTGACT 120
 DB 21 IleAlaAlaGluAspValMetIleValSerAlaSerGlyTyrGluLysLeuThr 40
 QY 121 AACGACGCGCCAGTGTCTGTGATTAGCCAGAGAGAAATTCAGTCCAGCCAGTACCAC 180
 DB 41 AsnAlaAlaSerValSerValIleSerGlnGluLeuGlnSerSerGlnTyrHis 60
 QY 181 GATCTGGCGGAGCTCTGAGATCAGTACAGGCTGGATGTTGAAAGTGGTACGGTAAA 240
 DB 61 AspLeuAlaGluAlaLeuArgSerValGluGlyValAspValGluSerGlyThrGlyLys 80
 QY 241 ACCGAGGCGTGAATCAGATCCGAGAAATGCCAGCCAGTTACACCTGATCTGATT 300
 DB 81 ThrGlyGlyLeuGluIleSerIleArgGlyMetProAlaSerTyrThrLeuIleLeu 100
 QY 301 GATGCTTCGACGGCGGAGCAGTACGCTGACTCCCAAGCTTTTCGCCATGAAT 360
 DB 101 AspGlyValArgGlnGlySerSerAspValThrProAsnGlyPheSerAlaMetAsn 120
 QY 361 ACCGAGTTCATCCCTCTGCGCCCATTTGAGCGTATTGAGGTTATCAGGGGCGGATG 420
 DB 121 ThrGlyPheMetProProLeuAlaIleGluArgGlyValIleArgGlyProMet 140
 QY 421 TCACACTGTATGGCTCTGATCGATGGCGGTGGTGGTGAATATCATACCAAGAGAAAT 480
 DB 141 SerThrLeuTyrGlySerAspAlaMetGlyGlyValValAlaSerIleThrArgLysAsn 160
 QY 481 GCAGACAAATGCTCTCTCCCTCAATGCAGGCTGAATCTGCAGGAAAGCAAAATGG 540
 DB 161 AlaAspLysTrpLeuSerSerValAsnAlaGlyLeuAsnLeuGlnSerAsnLysTrp 180
 QY 541 GGTAAACAGCAGCAGTTAATTTCTGGACAGATGTCCTCCCTTGGGATGATTCTGTACG 600
 DB 181 GlyAsnSerSerGlnPheAsnPheTrpSerSerGlyProLeuValAspSerValSer 200
 QY 601 CTGCGAGTACGGGTAGCACACACAGCGTCAGGCTTCATCGTCCATCATCTGAGCGAT 660
 DB 201 LeuGlnValArgGlySerThrGlnGlnArgGlnGlySerSerValThrSerLeuSerAsp 220
 QY 661 ACAGCAGGACCGGTATCTTCATCCACGGAGTCAACAATTAATCTTGGTGCAGCT 720
 DB 221 ThrAlaAlaThrArgIleProTyrProThrGluSerGlnAsnTyrAsnLeuGlyAlaArg 240
 QY 721 CTTGACTGAAGCGCTCGGACGAGATGCTCTGGTTGTATATGATACCAACCGCGGAG 780
 DB 241 LeuAspTrpIleAlaSerGlnGluAspValLeuTrpPheAspMetAspThrThrArgGln 260
 QY 781 CGTTATGATAACCGGGATGGCAACTGGGGAGCTCTCAGCGGGGATATGACCGGACCCCTG 840
 DB 261 ArgTyrAspAsnArgAspGlyGlnLeuGlySerLeuThrGlyGlyTyrAspArgThrLeu 280

QY 841 CGCTATGACGAAACAAAATTTCACTGCTGCTATGATCATCTTTTTCACCTTCGAAACATGG 900
 DB 281 ArgTyrGluArgAsnLysIleSerAlaGlyTyrAspHisThrPheThrPheGlyThrTrp 300
 QY 901 AAATCGTATCTGAACCTGGAAACGAGACAGAAATAAAGTCTGAGCTGTACCGAGTGA 960
 DB 301 LysSerTyrLeuAsnTrpAsnGluThrGluAsnLysGlyArgGluLeuValArgSerVal 320
 QY 961 CTGAAGCGCGACAAATGGGGCTTCCGCTCAGCCGCGAGCTTAAGAAATCGAACCTTT 1020
 DB 321 LeuLysArgAspLysTrpGlyLeuAlaGlyGlnProArgGluLeuLysGluSerAsnLeu 340
 QY 1021 ATCTCAATTCATTACTGTTACCTCTGGGAGATCTCATCTGTTACCGTGGGGGCG 1080
 DB 341 IleLeuAsnSerLeuLeuLeuThrProLeuGlyGluSerHisLeuValThrValGlyGly 360
 QY 1081 GAGTTTCAGAGCTCGTCCATGAAAGACGAGTGTCTTTCGCCAGCACAGGTGAAACTTTC 1140
 DB 361 GluPheGlnSerSerSerMetLysAspGlyValValLeuAlaSerThrGlyLysThrPhe 380
 QY 1141 CGGACAGAAAGCTGCTGCTGTTTCTGAGGATGAGTGGCATCTCAGGATCAGTTCGG 1200
 DB 381 ArgGlnLysSerTrpSerValPheAlaGluAspGluTrpHisLeuThrAspAlaLeuAla 400
 QY 1201 CTGACTGCGGCGAGCGCTATGAAATCATGAGCAATTCGGGGGACACTTTCAGTCCGGCT 1260
 DB 401 LeuThrAlaGlySerArgTyrGluHisGluGlnPheGlyGlyHisPheSerProArg 420
 QY 1261 GCATATCTGCTGCGGATGTCGAGATGCTGAGCTGAGCTGAAAGCGGTGTGACACGGGA 1320
 DB 421 AlaTyrLeuValTrpAspValAlaAspAlaTrpThrLeuLysGlyGlyValThrGly 440
 QY 1321 TATAAGCCACCCAGATGGGCGAGCTACATAAAGGATAGTGTGTGTGTCGCCGCGAGGA 1380
 DB 441 TyrLysAlaProArgMetGlyGlnLeuHisLysGlyIleSerGlyValSerGlyGlnGly 460
 QY 1381 AAAACAAATCTACTTGGTAAACCCCGACCTGAAGCCGGAAGAGAGCGTCACTTATGAGCT 1440
 DB 461 LysThrAsnLeuLeuGlyAsnProAspLeuLysProGluGluSerValSerTyrGluAla 480
 QY 1441 GGGGTGTATACGATAAACCCCGCTGTAATGCCAATGTCCACAGTTTATGACTGAC 1500
 DB 481 GlyValTyrTyrAspAsnProAlaGlyLeuAsnAlaAsnValThrGlyPheMetThrAsp 500
 QY 1501 TTCTCCAAACAAATGCTCTCTTATTCATAAATGATAACACCAATAGCTATTAACACAGC 1560
 DB 501 PheSerAsnLysIleValSerTyrSerIleAsnAspAsnThrAsnSerTyrValAsnSer 520
 QY 1561 GGAAGGCGCGTTCAGCGGTGGAAATTTGCGGCACATTCGCCCTGCTGGTCAGAGGAT 1620
 DB 521 GlyLysAlaArgLeuHisGlyValGluPheAlaGlyThrLeuProLeuTrpSerGluAsp 540
 QY 1621 GTCACTGTCTGCTGAATTTACACTCGACCCGAAGTGAACCAACGATGCTGATTAACAAA 1680
 DB 541 ValThrLeuSerLeuAsnTyrThrTrpThrArgSerGluGlnArgAspGlyAspAsnLys 560
 QY 1681 GGTGCGCGCTGAGTTATACCCCTGAACACATGGTGAATGCCAACTGAACCTGGCGAGTC 1740
 DB 561 GlyAlaProLeuSerTyrThrProGluHisMetValAsnAlaLysLeuAsnTrpGlnIle 580
 QY 1741 ACCGAGAGTGGCATCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1800
 DB 581 ThrGluGluValAlaSerTrpLeuGlyAlaArgTyrArgGlyLysThrProArgPheThr 600
 QY 1801 CAGAATTTATTCGTACTGAGCGCTGTACAGAAAGAAAGTGTATGATGAGAAAGAGAGATAC 1860
 DB 601 GlnAsnTyrSerSerLeuSerAlaValGlnLysLysValTyrAspGluLysGlyGluTyr 620
 QY 1861 CTGAAGCTGGAGCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1920
 DB 621 LeuLysAlaAlaTrpThrValValAspAlaGlyLeuSerTrpLysMetThrAspAlaLeuThr 640
 QY 1921 CTGAATGCTCGGTGAATTAACCTGCTCAACAAAGGATTACAGTGACGTGACGCTGTACAGT 1980

```

Db 641 LeuAsnAlaAlaValAsnAsnLeuLeuAsnLysAspTyrSerAspValSerLeuTyrSer 660
QY 1981 GCCGGTAAGACTACGCTGTATGCGGTGATTCTCCAGACGGGATCATCAACACAGGA 2040
Db 661 AlaGlyLysSerThrLeuTyrAlaGlyAspTyrPheGlnThrGlySerSerThrThrGly 680
QY 2041 TATGTGATACCTGAGCGAAATTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2088
Db 681 TyrValIleProGluArgAsnTyrTrpMetSerLeuAsnTyrGlnPhe 696

RESULT 2
AAR75366
ID AAR75366 standard; protein; 696 AA.
AC AAR75366;
XX
DT 13-MAY-1996 (first entry)
DE Adhesin.
XX
KW adhesin; plasmid pear; vector; vaccine; intestine colonisation.
XX
OS Escherichia coli.
XX
Key Location/Qualifiers
FH /note= "deduced residue from nucleotide sequence is His,
FT this differs from Seq ID5, residue is Gly"
FT
XX
PN WO9600233-A1.
XX
PD 04-JAN-1996.
XX
PF 07-JUN-1995; 95WO-US006994.
XX
PR 24-JUN-1994; 94US-00265714.
XX
PA (CHIL-) CHILDREN'S HOSPITAL & MEDICAL CENT.
PA (UNIW ) UNIV WASHINGTON.
PA (UNIW ) UNIV WASHINGTON STATE RES FOUND.
XX
PI Tarr PI, Bilge SS, Besser TE, Vary JC;
XX
DR WPI; 1996-068826/07.
DR N-PSDB; AAT10105.
XX
PT Chromosomal DNA from E. coli O157:H7 encoding epithelial adhesin -
PT isolated on plasmid pSC (overlap), for use as a vaccine to mediate
PT bacterial colonisation of bovine intestine.
XX
PS Disclosure; Page 31-32; 42pp; English.
XX
CC A candidate adhesin (AAR75366) was identified that is a homologue of the
CC IrqA protein of Vibrio cholerae. The adhesin enables Escherichia coli
CC O157:H7, an antibiotic-resistant, virulent and common food-borne
CC pathogen, to adhere to epithelial cells. Recombinant adhesin was obtd. by
CC expression of the encoding sequence (see AAT10105) in E. coli HB101
CC (pear). The adhesin can be used as a vaccine for immunisation of cattle
CC against disease or colonisation of mucosal surfaces by O157:H7, thus
CC increasing the safety of food derived from cattle
XX
SQ Sequence 696 AA;

```

```

Alignment Scores:
Pred. No.: 0 Length: 696
Score: 3628.00 Matches: 695
Percent Similarity: 99.86% Conservative: 0
Best Local Similarity: 99.86% Mismatches: 1
Query Match: 97.21% Indels: 0
DE: 2 Gaps: 0

```

US-10-625-972-4 (1-2091) x AAR75366 (1-696)

```

QY 1 ATCCGAATACCACTCTCGCTTCCTAGTCAATCCCTCTCGGATTTTCACCGACGAC 60
Db 1 MetArgIleThrLeuAlaSerValIleProCysLeuGlyPheSerAlaSerSer 20
QY 61 ATAGTGTCTGCAGAGATGTGATGATTGTCTCGGCATCCGGCTATGAGAAAAGTGACT 120
Db 21 IleAlaAlaAlaGluAspValMetIleValSerAlaSerGlyTyrGluLysLeuThr 40
QY 121 AACGACGCCGCGAGTGTCTCTGATTAGCCAGGAGGAATTCAGTCCAGCCAGTACCAC 180
Db 41 AsnAlaAlaAlaSerValSerValIleSerGlnGluLeuGlnSerSerGlnTyrHis 60
QY 181 GATCTGGCGGAGCTCTGAGATCAGTAGAGGTGTGGATGTTGAAAGTGGTACCGGTAAA 240
Db 61 AspLeuAlaAlaGluAlaLeuArgSerValGluGlyValAspValGluSerGlyThrGlyLys 80
QY 241 ACCGAGGCGCTGGAATCAGCATCCGAGGAATCCAGCAGGTTACACGCTGATACGTATT 300
Db 81 ThrGlyGlyLeuGluIleSerIleArgGlyMetProAlaSerTyrThrLeuIleLeuIle 100
QY 301 GATGTGTTCGTACGGGGAGCAGTACGCTGCTCCCAACGGTTCCTCCCATGAAT 360
Db 101 AspGlyValArgGlnGlyGlySerSerAspValThrProAsnGlyPheSerAlaMetAsn 120
QY 361 ACCGGGTTTCATCCCTCTGCGCCGCTATTGAGCGTATTGAGGTTATCAGGGGCGCGATG 420
Db 121 ThrGlyPheMetProProLeuAlaAlaIleGluAlaGlyValIleArgGlyProMet 140
QY 421 TCCACACTGTATGCTCTGATCGGATGGCGGTGGTGAATATCATACCAAGAAGAT 480
Db 141 SerThrLeuTyrGlySerAspAlaMetGlyGlyValValAsnIleIleThrArgLysAsn 160
QY 481 GCAGACAAATGCTCTCTTCCTCAATGCAGGCTGGAATCTGCAGGAAACACAAATGG 540
Db 161 AlaAspLysTrpLeuSerSerValAsnAlaGlyLeuAsnLeuGlnGluSerAsnLysTrp 180
QY 541 GGTAAACAGCAGCCAGTTAAATTTCTGGAGCAGTGTCCCTTGTGGATGATTCTCTCAGC 600
Db 181 GlyAsnSerSerGlnPheAsnPheTrpSerSerGlyProLeuValAspSerValSer 200
QY 601 CTGCAGGTACGGGTAGCACACACAGCGTCAAGGTTTCATCGTCCACATCCTCAGAGCAT 660
Db 201 LeuGlnValArgGlySerThrGlnGlnArgGlnGlySerSerValThrSerLeuSerAsp 220
QY 661 ACAGCAGCAGCGGTATTCCTTATCCCGAGGTACACAGAATTATAATCTTGGTCCACGT 720
Db 221 ThrAlaGlyThrArgIleProTyrProThrGluSerGlnAsnTyrAsnLeuGlyAlaArg 240
QY 721 CTGTGCTGGAAGCGCTCGGACGAGATGTCTCTGTTGTATGATGATACACCCCGGCGAG 780
Db 241 LeuAspTrpLysAlaSerGluGlnAspValLeuTrpPheAspMetAspThrThrArgGln 260
QY 781 CGTTATGATAACCGGGATGGCACTGGGGAGTCTGACGGGGGATATGACCGGACCCCTG 840
Db 261 ArgTyrAspAsnArgAspGlyGlnLeuGlySerLeuThrGlyTyrAspArgThrLeu 280
QY 841 CCCTATGACGCGAAACAAATTCACGCTGCTGATGATCATCTTCACCTTCGGAACATGG 900
Db 281 ArgTyrGluArgAsnLysIleSerAlaGlyTyrAspGlyThrPheThrPheGlyThrTrp 300
QY 901 AAATCGTATCTCAACTGGAAACGAGACAGAAATAAAGTCTGAGCTTGTACGCACTGTA 960
Db 301 LysSerTyrLeuAsnTrpAsnGlnThrGluAsnLysGlyArgGluLeuValArgSerVal 320
QY 961 CTGAACGCGCAAAATGGGGCTTCGGGTACGCGCGGGAGCTTAAAGAAATCGAACCTT 1020
Db 321 LeuLysArgAspLysTrpGlyLeuAlaGlyGlnProArgGluLeuLysGluSerAsnLeu 340
QY 1021 ATCCCTGAATTCATTACTGCTTACCCCTCTGGAGATCTCATCTGTTACCGTGGGGGC 1080
Db 341 IleLeuAsnSerLeuLeuLeuThrProLeuGlyGluSerHisLeuValThrValGlyGly 360

```


1081	QY	GAGTTTCAGAGCTCGTCCATGAAAGCAGGAGTTGTCTTCCTGCCAGCAGACAGGTGAACCTTTC	1144
361	Db	GlupheGlnSerSerMetLysAspGlyValValLeuAlaSerThrGlyGluThrPhe	380
1141	QY	CGGCAGAAAAGCTGTCGGGTATTTCGTGAGGATGAGTGGCATCTCAGGATGACACTTGC	1200
381	Db	ArgGlnLysSerTrpSerValPheAlaGluAspGluTrpHisLeuThrAspAlaLeuAla	400
1201	QY	CTGACTCGGGCAGCCGCTATGAACATCATGAGCAATTCGGGGAGCACTTCAGTCCGCGT	1260
401	Db	LeuThrAlaGlySerArgTyrGluHisGluGlnPheGlyGlyHisPheSerProArg	420
1261	QY	GCATATCTGTCCTGGGATGTCGACAGATCGCTCGAGCGCTGAAGAGCGGTGTACCCACGGGA	1320
421	Db	AlaTyrLeuValTrpAspValAlaAspAlaIatrpThrLeuLysGlyGlyValThrTrpGly	440
1321	QY	TATAAGCCACCCAGATGGCGCAGCTACATAAAGGGATTAGTGTGTGTCTCGGGCAGGGA	1380
441	Db	TyrLysAlaIatrpArgMetGlyGlnLeuHisLysGlyLleSerGlyValSerGlyGlnGly	460
1381	QY	AAAAACAATCTACTTGTAAACCCGACCTGAAGCCGGAAGAGAGCGTCAGTTATGAGCCT	1440
461	Db	LysThrAsnLeuLeuGlyAsnProAspLeuLysProGluGluSerValSerTyrGluAla	480
1441	QY	GGGTGTATTACAGATAACCCCGCGCTCTGAATGCCAATGTCAAGTTTATGACTGCAC	1500
481	Db	GlyValTyrTyrAspAsnProAlaGlyLeuAsnAlaAsnValThrGlyPheMetThrAsp	500
1501	QY	TTCTCCAAAGATGTCCTTATTCATAAATAGATAACACCATAGCTATGTTAAACAGC	1560
501	Db	PheSerAsnLysIleValSerTyrSerLleAsnAspAsnThrAsnSerTyrValAsnSer	520
1561	QY	CGAAGCCCGGTGTCACCGTGGAAATTGCGGCACATTGCCGCTGTGTGTCAGAGCAT	1620
521	Db	GlyLysAlaArgLeuHisGlyValGluPheAlaGlyThrLeuProLeuTrpSerGluAsp	540
1621	QY	GTCACGCTGTCACGTGAATTACACCTCGACCCCGAAGTGAAACAACGTGATGTATAACAA	1680
541	Db	ValThrLeuSerLeuAsnTyrThrTrpThrArgSerGluGlnArgAspGlyAspAsnLys	560
1681	QY	GCTCGCCGCTGAGTTATACCCCTGAACACACATCGTGAATGCGAACTGAACCTGGCAGATC	1740
561	Db	GlyAlaProLeuSerTyrThrProGluHisMetValAsnAlaLysLeuAsnTrpGlnIle	580
1741	QY	ACCCAAGAGGTGGCATCATGCTGGGTGGCCGCTTATCGCGGGAACACCACTTTCACC	1800
581	Db	ThrGluGluValAlaSerTrpLeuGlyAlaArgTyrArgGlyLysThrProArgPheThr	600
1801	QY	CAGATTATTTCGTACATGACCGCTGTACAGAGAAGTGTATGATGAGAAGAGGATATC	1860
601	Db	GlnAsnTyrSerSerLeuSerAlaValGlnLysLysValTyrAspGluLysGlyGluTyr	620
1861	QY	CTGAAGCCTCGACGGTGGGATGCAGCTCTGCTGTGAAGATGACGGATGCCCTCAGC	1920
621	Db	LeuLysAlaTrpThrValValAspAlaGlyLeuSerTrpLysMetThrAspAlaLeuThr	640
1921	QY	CTGAATGCTCGGTGAATAACTGCTCAACAGAGATTACGTGACGTGAGCCTGTACAGT	1980
641	Db	LeuAsnAlaAlaValAsnAsnLeuLeuAsnLysAspTyrSerAspValSerLeuTyrSer	660
1981	QY	GCCGTTAAGAGTACGCTGTATGCGGTGATTCTTCTCCAGACGGGATCATCAACAACAGGA	2040
661	Db	AlaGlyLysSerThrLeuTyrAlaGlyAspTyrPheGlnThrGlySerSerThrThrGly	680
2041	QY	TATGTGATACCTGACGGAATTACTGGATGTCGCTGAACATTCAGTTC	2088
681	Db	TyrValIleProGluAArgAsnTyrTrpMetSerLeuAsnTyrGlnPhe	696
RESULT 3			
ABM67028			
ID	ABM67028	standard; protein; 660 AA.	
XX			
AC	ABM67028;		

Db SerAsnAspThrIleLeuValThrAlaAlaGlyPheGlnGlnLysIleGluAspAla 47
 QY 127 GCCGCCAGTGTTCGTGATTAAGCAGAGAAATGTCAGTCCAGCCAGTACACGATCTG 186
 Db ProAlaSerIleSerValSerArgGluGlnLeuGluThrLysAlaTyrArgAspVal 67
 QY 187 GCGAGGCTCTGACATCAGTAGAGGGGTGGTATCTTCAAAGTGTGACGGTAAACCCGGA 246
 Db ThrAspAlaLeuLysAspValProGlyValValVal-----ThrGlyGlyAlaSer 84
 QY 247 GGGTGGAAATCAGATCCCGAGGAATCCAGCCAGTATACAGCTGATGATGATGATGAT 306
 Db SerSerAspIleSerIleArgGlyMetAlaAlaGlnTyrThrMetIleLeuValAspGly 104
 QY 307 GTTCGTGAGGCGGAGCAGTACGTGACTCCCAACCGTTCCTGTCATGAATACCGGG 366
 Db LysArgValAspThrArgGlyThrArgProAsnArgAspAsnSerGlyIleGluGlnGly 124
 QY 367 TTCATGCCCTCTGGCCGCAATGACGCTATTGAGTTATCAGGGGGCCGATGTCACCA 426
 Db TrpLeuProLeuAlaAlaIleGluArgIleGluValValArgGlyProMetSerSer 144
 QY 427 CTGTATGGCTGTGATCGATGGGGGTGGTGTGATATCATATACCAAGAAAGATGCAGAC 486
 Db LeuTyrGlySerAspAlaMetGlyGlyValIleAsnIleIleThrArgLysAlaGlnLys 164
 QY 487 AAATGGCTCTCTCCGTCAATGTCAGGGCTGAATCTGCAGGAAGCAACAAATGGGTAAC 546
 Db GluTrpLysThrSerPheArgAlaAspGlyThrLeuGlnGluArgSerLysSerGlyAsn 184
 QY 547 AGCAGCCAGTTTAACTTCTGGAGCAGTGGTCCCTTGTGGATGATTTCTGCACGCTGCAG 606
 Db SerHisGlnSerSerValTyrThrSerGlyProIleIleAspGlyLeuLeuGlyLeuLys 204
 QY 607 GTACGCGGTACACACACAGCAGTCCAGGGTTCATCGGTTCATCAGTACGAGCGATACAGCA 666
 Db ValSerGlyLeuTyrSerHisArgAsnGluAspLysPhe----- 217
 QY 667 GGCACGGGTATTCCTTATCCACAGAGTCACAGAAATTAATCTGTGTGTCACGCTTTCAC 726
 Db -----IleGlyGlyPheAsnLysGlnGluMetArgAsnGlySerAlaThrPheSer 234
 QY 727 TGGAAAGCGTCGGAGCAGGATGCTCTGTGTTGATATGATGATACACCCGCGCAGCTTAT 786
 Db PheThrProAspGluGlnAsnThr-----PheAspPheGluAlaGlyArgTyrGluGln 252
 QY 787 GATAACCGGATGGCAACTCGGGAGTCTCAGCGGGGATATGACCGGAC----- 837
 Db Asp---ArgAspSerThrValGlyLysThrLysThrSerSerTrpArgSerProGlyAsn 271
 QY 838 -----CTGCCCTATAGCGAAACAAATTTACGCTGGCTATGATCATCTTTCACCTTC 891
 Db SerGluSerArgTyrLysArgAsnAsnTyrSerIleThrHisAsnGlyValTyrAspPhe 291
 QY 892 GGAACATGGAATCGTATCTCACTGGAACAGAGACAGAAATAAAGTCTGTGAGCTTGTGA 951
 Db GlyThrSerThrThrTyrPheGluArgAspGluSerArgAsnProGly----- 307
 QY 952 CGCAGTGTACTGAAGCGCGCAAAATGGGGGCTTCGCGGTACGCGCGGAGCTTAAAGAA 1011
 Db -----ArgAspMetLysTyr 312
 QY 1012 TCGAACCTTATCCTGAATTCATTACTGTTACCTCTGGAGAAATCTCATCTGTTTACG 1071
 Db TyrAspAsnLeuPheAsnThrHisThrValPheMetLeu---AspAspHisSerLeuSer 331
 QY 1072 GTGGGGGCGAGTTTCAGAGCTCGTCCATGAAGAC---GGAGTTGTCTTTCGCGAGCACA 1128
 Db IleGlyGlyGlnTyrArgTyrGluAspLeuArgAspGluGlyAsnGlnLeuAlaSerAla 351
 QY 1129 GGT-----GAAACTTTCGCGCAGAAAGCTGGTATTTGCTGAGGATGAGTGCAT 1182
 Db LysAsnIleAsnLysLeuAsnArgTrpSerTrpAlaLeuPheAlaGluAspGluTrpGln 371

QY 1183 CTCACGGATGCACCTTGCCTGACTCGCGGACCGCTATGACATCATGACCAATTCGGG 1242
 Db MetThrAsnAspPheAlaLeuThrGlyGlyIleArgMetAspLysAspGluAsnPheGly 391
 QY 1243 GGACACTTCAGTCCCGCTGCATATCTGCTGGGATGTGGCAGATGCCCTGGACGCTGAAA 1302
 Db AspHisTrpThrProArgLeuTyrGlyValTrpHisValAspGluGlnTrpThrIleLys 411
 QY 1303 GCGGTGTGACACCGGATATAAGCACCAGCAATGGGGCTACATATAAAGGATAGT 1362
 Db GlyGlyValSerThrGlyTyrArgSerProGluLeuArgGlnAlaAlaSerAsnTrpGly 431
 QY 1363 GGTGTGTCGCGGACGCGA-----AAAACAAATCTACTTGTGTAAACCCGACCTGAAGCG 1416
 Db GlnIleThrGlyGlyGlyAspProAlaIleIleValGlyAsnProAspLeuLysPro 451
 QY 1417 GAAAGAGCGTCAGTTATGAGCTGGGTGTATTAGATTAACCCCGCGCTGTAATGCC 1476
 Db GluLysSerPheAsnGlnIleGlyValIleTrpAsnAsnArgAspAsnPheAsnVal 471
 QY 1477 AATGTCACAGGTTTATGACTGACTCTCTCAACAGATGTCTCTAT----- 1524
 Db GlyLeuThrIlePheAsnThrGluPheLysAspLysIleThrGluLeuArgArgCysThr 491
 QY 1525 -----TCCATAAATGATAACACC----- 1542
 Db AspProSerGlyLysSerGlnGlyLysCysValIleAspGlyThrAlaTyrLysPheIle 511
 QY 1543 AATAGCTATGATAACAGCGGAAAGCCCGGTGTCAGCTGTGGAATTTCCGCGCACATG 1602
 Db SerArgAspValAsnIleAspLysAlaAsnMetArgGlyValGlu-----AlaThrMet 529
 QY 1603 CCGCTGTGTGTC-----GAGGATGTCAGCTGTCTCACTGAATTCACCTGGACCCGAAAT 1656
 Db Asn---TipAspIleValAspAsnTrpValLeuAlaAlaAsnTyrThrTyrThrAspSer 548
 QY 1657 GAAACAGCTGATGTCATTAACAAAGTGGCGCTGAGTTATACCTCAACACACATGGTG 1716
 Db GluGlnLysSerGlyAsnPheLysGlyLysProLeuAsnLysMetProLysHisMetAla 568
 QY 1717 AATGCGAAACTCAACTGGCAGATCACCGAAGAGTGGCATCATGCTGGTGGCTGCTTAT 1776
 Db AsnAlaThrLeuAsnTrpGlnAlaSerProGluMetGlnAlaTrpThrArgValAsnTyr 588
 QY 1777 CCGCGGAAACACACAGCTTTCACCGAGAATTAATCGTCAGCGCTGTACAGAGAAA 1836
 Db ArgGlyLys-----ThrSerAsnTyrLeuSerArgThrSerMet----- 601
 QY 1837 GTGTATGATGAGAAAGGAGAAATACCTGAAAGCTCGACGGTGTGATCGAGTCTGTGCG 1896
 Db -----GlyThrGlyThrProSerTyrThrPheValAspIleGlyValSer 616
 QY 1897 TGGAAAGATGACGGATGCCCTGACGCTGAATGTCGGGTGAATAACCTGCTCAACAGGAT 1956
 Db TyrTyrLeuAlaGluGlyLeuArgLeuValGlyGlyValTyrAsnValPheGluLys--- 635
 QY 1957 TACAGTACGCTGAGCCTGTACAGTCCGCGTAGAGTACGCTGTATGCCGCTGATTACTTC 2016
 Db ----- 635
 QY 2017 CAGACGGGATCATCAACACAGGATATGTATACCTGAGCGAAATTAATGATGCTGCTG 2076
 Db GlnValThrGluAspThrHisAspIleValLeuAspGlyArgTyrAsnIleGlyIle 655
 QY 2077 AACTATCAGTTC 2088
 Db AsnTyrAspPhe 659
 RESULT 4
 ID ABB52836 standard; protein; 682 AA.
 XX

AC ABB52836;
 XX 11-FEB-2002 (first entry)
 XX Escherichia coli polypeptide SEQ ID NO 1077.
 DE
 XX Escherichia coli; B2/D+A-; antiinflammatory; antibacterial;
 XX immunosuppressive; extra-intestinal infection; phylogeny; meningitis;
 XX systemic infection; non-diarrhoeal infection; septicaemia;
 XX pyelonephritis; antibiotic resistance.
 XX
 OS Escherichia coli.
 XX
 XX WO200166572-A2.
 PN
 XX 13-SEP-2001.
 PD
 XX 12-MAR-2001; 2001WO-EP003445.
 XX
 XX 10-MAR-2000; 2000PR-00003145.
 PR
 XX 02-FEB-2001; 2001PR-00001449.
 XX
 XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 XX
 XX Bingen E, Bonacorsi S, Clermont O, Nassif X, Tinsley C;
 PI
 XX WPI; 2001-550253/61.
 DR
 XX
 XX A library of DNA fragments of Escherichia coli strains for the phylogenetic
 PT determination of a given strain comprises polynucleotides of nature B2/D+
 PT A-.
 XX
 XX Example 6; Fig 6; 646pp; English.
 PS
 XX The invention relates to a library of DNA fragments of Escherichia coli
 CC strains comprising polynucleotides (AB48577-AB488729 and AB489533) and
 CC encoded proteins (AB52459-AB52919 and AB52954-AB53094) of nature
 CC B2/D+A-. The polynucleotides have potential antiinflammatory,
 CC antibacterial and immunosuppressive activity as part of pharmaceutical
 CC compositions used to treat, palliate or prevent extra-intestinal E. coli
 CC infections. The polypeptides are useful for determining the phylogenetic
 CC group of a given E. coli strain. These polypeptides can detect and treat
 CC an undesired development of E. coli, particularly an extra-intestinal
 CC infection that include systemic and non-diarrhoeal infections such as
 CC septicaemia, pyelonephritis and meningitis this is particularly
 CC advantageous as bacterial resistance is increasing with the more frequent
 CC use of broad spectrum antibiotics
 XX
 XX Sequence 682 AA;
 SQ

Alignment Scores:
 Pred. No.: 2,896-81 Length: 682
 Score: 972.00 Matches: 244
 Percent Similarity: 51.08% Conservative: 135
 Best Local Similarity: 32.88% Mismatches: 245
 Query Match: 26.05% Indels: 118
 DB: 4 Gaps: 23

US-10-625-972-4 (1-2091) x ABB52836 (1-682)

QY 7 ATACACACTCGCTCGGTAGTATTCCTCTCGGATTTTCAGCCAGCAGCATAGCT 66
 DB 11 lLeAlaValMeCySerAlaValIleSerSer---GlyTyrAlaSerSerAsp----- 27
 QY 67 GCTGCAGAGGATGTGATGATTCCTCGGCATCCGCTATGAGAAAGCTGACTAAGCCA 126
 DB 28 LysLysGluAspThrLeuValValThrAlaSerGlyPheThrGlnGlnLeuArgAsnAla 47
 QY 127 GCCGCCAGGTTTCTGTGATTAACGAGGAGGATTCAGCTCCAGCCAGTACCAGATCTG 186
 DB 48 ProAlaSerValSerValIleThrSerGluGlnLeuLysLysProValSerAspLeu 67
 QY 187 GCGGAGGCTCTGAGATCAGTAGAGGGTGTGGATGTTGAAGTGGTACGGGTAAACCCGA 246

DB 68 ValAspAlaValLysAspValGluGlyIleSerIleThrGlyGlyAsnGluLysPro--- 86
 QY 247 GGGCTGGAATCAGCATCCGAGGAATGCCAGCAGTATACCGCTGATACTGATTGATGTT 306
 DB ---AspIleSerIleArgGlyLeuSerGlyAspTyrThrLeuIleLeuValAspGly 104
 QY 307 GTTCGTGAGGGGGAAGCAGTACGTGACTCCCAACGGTTTTCGCCATGAATACCGGG 366
 DB 105 ArgArgGlnSerGly---ArgGluSerArgProAsnGlySerGlyGlyPheGluAlaGly 123
 QY 367 TTCATGCCCTCTCGCCCATTCAGCGTATGTAGGTTATCAGGGGCGCGATGTCCACA 426
 DB 124 PheIleProValGluAlaIleGluArgIleGluValIleArgGlyProMetSerSer 143
 QY 427 CTGTATGGCTCTGATCGCATGGCGGTGGTGGTGAATATCATATACCATCAGAAAG--- 483
 DB 144 LeuTyrGlySerAspAlaIleGlyValIleAsnIleIleThrLysProValAsnAsn 163
 QY 484 GACAAATGGCTCTCTCCGCTCAATGCGGGCTGAATCTGCAGAAAGCAACAATAGGGGT 543
 DB 164 GlnThrTyrAspGlyValLeuGlyLeuGlyIleIleGlnGluHisGlyLysPheGly 183
 QY 544 AACAGCAGCAGTTAAATTTCTGGAGCAGTGTCCCTTGTGGATGATCTCTGACCGTG 603
 DB 184 AsnSerThrThrAsnAspPheTyrLeuSerGlyProLeuIleLysAspLysLeuGlyLeu 203
 QY 604 CAGGTACGGGTAGCACACACAGCGTCTAGGGTTCATCGTCCATCATCTAGCGCATACA 663
 DB 204 GlnLeuTyrGlyGlyMetAsnTyrArgLysGluAspSerIleSerGln----- 219
 QY 664 GCAGGCACCGCTATTCTCTTATCCACGGAGTCCAGAAATTAATATCTTGGTCCAGCTTT 723
 DB 220 ---GlyThr-----ProAlaLysAspAsnLysAsnIleThrAlaThrLeu 233
 QY 724 GACTGGAAGCGCTCGAGCAGGAGTGTCTCTGTTTGTATATGATATACACCCGCGAGGT 783
 DB 234 GlnPheThrProThrGluSerGlnLysPheValPheGluTyr----- 247
 QY 784 TATGATAACCGGATGGGCAACTCTGCGGGGAGTCTGACGGGGGATATGACCGGACCTGGCG 843
 DB 248 ---GlyLysAsnAsnGlnValHisThrLeuThrProGlyGluSerLeuAspAlaTrp 265
 QY 844 TATGACGGAACAAATTTACGTGCTGCTATGATCATCTTCACCTTCGGAACATGGAA 903
 DB 266 ThrMetArgGlyAsnLeuLysGlnProAsnSerLysArgGluThrHisAsnSerArgSer 285
 QY 904 TCGTATCTG---AACTGGAAAC-----GAGACAGAAAAATAAA 936
 DB 286 HisTrpValAlaAlaTrpAsnAlaGlnGlyGluIleLeuHisProGluIleAlaValTyr 305
 QY 937 GGTCTGAGCTTGTACGAGTGTACTG-----AGCGGCACAAA-----TGGGGG 981
 DB 306 GlnGluLysValIleArgGluValLysSerGlyLysLysAspLysTyrAsnHisTrpAsp 325
 QY 982 CTTGCGCGTCCAGCCGCGAGCTTAAGGAATCGAATCTTCTGAAATTCATTACTGCTT 1041
 DB 326 LeuAsnTyrGluSerArgLysProGluIleThrAsnThrIleLeuAspAlaLysValThr 345
 QY 1042 ACCCTCTGGGAGATCTCATCTGTTACGTGGGGGCGAGTTTCAGAGCTCGTCCATG 1101
 DB 346 AlaPheLeuProGlu---AsnValLeuThrIleGlyGlyGlnPheGlnHisAlaGluLeu 364
 QY 1102 AAGACGGAGTTCCTCTCCAGCACAGGTGAA-----ACT 1137
 DB 365 ArgAspAsp-----SerAlaThrGlyLysLysThrGluThrGlnSerValSer 381
 QY 1138 TTCGCGCAGAAAAGCTGTGCGGTATTTCGTAGGATGATGCGCATCTCTCAGGATCATT 1197
 DB 382 IleLysGlnLys-----AlaValPheIleGluAsnGluTyrAlaAlaThrAspSerLeu 399
 QY 1198 GCGCTACCTCGCGGAGCGCTATGACATCATGACATTCGGGGGACACTTCAGTCCG 1257

115	Db	115	IleArgGlyMetGlyProGluAsnThrLeuValAspGlyLysProIleAsnSer	134
307	Qy	307	-----GTTGTCAGGGCGGAAGCAGTGCAGTGCTCCCAACGGTTTTCTGCCATG	357
135	Db	135	ArgAsnSerValAlaGlyGlyTrpLysGlyAlaArgAspThrArgGlyAspSerAsn---	153
358	Qy	358	AATACCGGGTTTCATGCCCCCTCTGGCCGCCATTGAGCGTATTGAGGTTATCAGGGGCGCG	417
154	Db	154	-----TtpValProAlaGluAlaIleGluSerLleGluValLeuArgGlyPro	169
418	Qy	418	ATGTCCACACTGTATGGCTCTCATCGCATGGCGGTGTGGTGAATATCATCACCAGAAG	477
170	Db	170	AlaAlaAlaAa-gtyGlySerGlyAlaAlaGlyGlyValAlaAsnIleIlethrLysLys	189
478	Qy	478	AATGCAGCAAAATGGCTCTCTTCCTCAATGCAGGGCTGCAATCTGCAGAAACGACAA	537
190	Db	190	ValThrAsnGluThrHisGlySerValGluIlePheTyrThrSerGlnProGluAspSerLys	209
538	Qy	538	TGGGTTACAGCAGCAGCAGCTTAAATTTCTGGAGCAGTGTCTCCCTTGGGATGATCTGTC	597
210	Db	210	GluGlySerSerAsnArgValGlyPheAsnValSerGlyProLeuIleLysAspValLeu	229
598	Qy	598	AGCTCTGAGGTACGGGTAGACACACACAGCGTCAAGGTTTCATCGTCT-----ACA	648
230	Db	230	SerTyrArgLeuTyrGlyAsnTyrAsnLysThrGluAlaAspAspValAspIleAsnLys	249
649	Qy	649	TCATGACCGATACAGCAGGCGCGGTATTCCTTATCCCGAGGTACACAGAATTATTAAT	708
250	Db	250	SerIleGlySerThrAlaAlaGlyArg-----GluGlyValLysAsnLysAsp	265
709	Qy	709	CTTGTCGACGCTTCATCTGGAAGCGTCGAGCAGGATGCTCTGCTTGTATGATGAT	768
266	Db	266	IleSerGlyValGLeuAlaTrpIleAlaThrAspGlnGlnThrValLeuLeuAspIleSer	285
769	Qy	769	ACCACCCGCGAG-----CGTTATGATAACCGG	795
286	Db	286	SerSerLysGlnGlyAsnIleTyr-SerGlyAspSerGlnLeuAsnAlaAsnAlaGluAla	305
796	Qy	796	GATGGCAACTGGGGAGTCTACCGGGGGATATGACCGGACCTCGCGTATGACGGAAC	855
306	Db	306	AspAlaIleLeuSerGlnLeuIleGlyLysGluThrAsnThrMet-----TyrArgAsp	323
856	Qy	856	AAAATTTTCAGCTGGCTATGATCATACTTTTCACCTTCGGAACATCGATATCTGAAC	915
324	Db	324	SerTyrAlaLeuThrHisGluGlyAspTrpSerTrpGlyLysSerLysLeuValAlaGln	343
916	Qy	916	TGGAAACGAGACAGAAAATAAAGCTGCTGAGCTGTACGCACTGTACGAACGGGACAA	975
344	Db	344	TyrAspLysThrHisAsn-----LysArgLeuPro	353
976	Qy	976	TGGGGCTTCGCGGTACGCCCGGGAGCTTAAGGAATCGAACCTT-----	1020
354	Db	354	GluGlyLeuAlaGlySerVal--GluGlyLysIleAsnAsnLeuAspAspLysAlaThr	372
1021	Qy	1021	---ATCCGTGAATCATTAATCTGCTTACCCCTCTGGAGAAATCTCATCTG-	1065
373	Db	373	SerArgLeuGluThrLeuArgPheAsn-----GlyGluAlaAsnIleProPheGluTyr	390
1066	Qy	1066	-----GTTACCGTGGGGCGAGTTTCAGAGCTCGTCCATCAAGACGGA	1110
391	Db	391	TyrLeuProGlnValLeuThrValGlyThrGluTrpValGluAspArgPheLysAspAsn	410
1111	Qy	1111	GTTGTCCTTGGC-----AGCACAGTGAAACTTTC-----	1140
411	Db	411	ValSerThrThrGlnGlyLysAspSerSerGlySerGlyTyrGlyAspGlnLeuAlaLys	430
1141	Qy	1141	-----CGCACAAAGCTGGTCTGGTATTTCGTAGGATGAGTGGCAT	1182
431	Db	431	GlyAspArgSerLysMetGluSerArgIleAlaSerAlaTyrIleGluAspAsnLeuLys	450
1183	Qy	1183	CTCAGGATGCACTGGCGTCACTCGCGGCGCCGCTTAAACATCATGAGCAATCGGG	1242

XX AC AAY72914;
XX DT 13-JUN-2001 (first entry)
XX DE E. coli ironNec extracytoplasmic protein fragment.
XX KW ironNec; extracytoplasmic protein; immunogen; vaccine; ET; UTI;
XX KW immunotherapy; extraintestinal infection; urinary tract infection;
XX KW meningitis; pneumonia; intra-abdominal infection; antibiotic.
XX OS Escherichia coli.
XX FH Key Location/Qualifiers
XX FT Peptide 1..24
XX FT /label= signal_peptide
XX FT Protein 25..725
XX FT /label= Mature_E. coli_ironNec_extracytoplasmic_protein
XX FT Misc-difference 120..121
XX FT /note= "Encoded by GTG CGT TAT"
XX FT Misc-difference 134..136
XX FT /note= "Encoded by AAC TGG"
XX FT WO200121636-A1.
XX PN 29-MAR-2001.
XX PD 22-SEP-2000; 2000WO-US026117.
XX PF 22-SEP-1999; 99US-0155621P.
XX PR (UNY) UNIV NEW YORK STATE RES FOUND.
XX FA Russo T, Carlino U;
XX FI WPI; 2001-244936/25.
XX FI N-PSDB; AAD03054.
XX DR Novel isolated ironNec polynucleotide from extraintestinal isolate of
XX PT Escherichia coli useful as vaccine for treating or preventing
XX PT extraintestinal infections caused by extraintestinal pathogenic
XX PT Escherichia coli.
XX PS Claim 9; Page 39-41; 44pp; English.
XX CC The invention relates to the identification of ironNec gene, from an
XX CC extraintestinal isolate of Escherichia coli. This gene is expressed in
XX CC increased amounts in human urine and is identified by transposon (mpHoA)
XX CC mutagenesis. ironNec gene encodes an extracytoplasmic protein. This gene
XX CC can be used as an immunogen in vaccine formulations. The recombinant
XX CC vector comprising nucleotide sequence encoding one or more antigenic
XX CC epitope of ironNec is useful for diagnostic and immunotherapeutic
XX CC purposes. The ironNec antigenic peptide is useful for treating or
XX CC preventing extraintestinal infections (Eis) caused by extra- intestinal
XX CC pathogenic E. coli (ExPEC). The Eis include urinary tract infection
XX CC (UTI), meningitis, intra-abdominal infection and pneumonia. The present
XX CC sequence is E. coli ironNec extracytoplasmic protein
XX SQ Sequence 725 AA;

Alignment Scores:
Pred. No.: 6,3e-47 Length: 725
Score: 604.50 Matches: 212
Percent Similarity: 43.88% Conservative: 136
Best Local Similarity: 26.73% Mismatches: 280
Query Match: 16.20% Indels: 165
DB: 4 Gaps: 34

US-10-625-972-4 (1-2091) x AAY72914 (1-725)
QY 1 ATGCCAAATACC---ACTCTGCTTCCTAGTCAATTCCTGTCGGATTTCAGCCAGC 57
DB 1 MetArgIleAsnLysIleLeuPsrLeuThrValLeuValGlyLeuAsn---Ser 19

QY 58 AGCATAGTCTGCTGCA-----GAGGATGTGATGATTGTCTCGGCA 96
DB 20 GlnValSerValAlaLysTyr-SerAspAspAspAspGluThrLeuValValGluAla 39
QY 97 TCCGGCTATGAAAGCTGACTAACCGACCGCCAGTCTTCTGTGATTAGCCAGAG 156
DB 40 ThrAla-----GluGlnValLeuLysGlnPnProGlyValSerValThrSerGlu 57
QY 157 GAATTCGATCCAGC---CAGTACCACGATCTGGCGGAGCTCTGAGATCAGTAGGGT 213
DB 58 AspIleLysLysThrProProValAsnAspLeuSerAspIleLeuLysMetProGly 77
QY 214 GTGGATGTT-----GAAAGTGTACGGGTAAACCCGA---GGCTGGAATCAGCATC 264
DB 78 ValAsnLeuThrGlyAsnSerAlaSerGlyThrArgGlyAsnAsnArgGlnIleAspIle 97
QY 265 CCAGGAATCCAGCCAGCATTACACGCTGATCTATGATTGATGTTGTCGTCAGCGCGGAAGC 324
DB 98 ArgGlyMetGlyProGluAsnThrLeuIleLeuIleAspGlyValProValThrSerArg 117
QY 325 AGTACGCTG-----ACTCCCAACGGTTTTTCTGCCATGAAATACC 363
DB 118 AsnSerValTyrSerTrpArgGlyGluArgAspThrArgGly-----AspThr 133
QY 364 GGGTTCATGCCCTCTGCGCCCATTCAGCGCTATTGAGGTTATCAGGGGCGCGATGCC 423
DB 134 AsnArgTrpValProProGluGlnValGluArgIleGluValIleArgGlyProAlaAla 153
QY 424 ACACCTGTATGGCTCTGATGCGGATGCGGCTGGTGAATATATCATCAGGAAGAATGCA 483
DB 154 AlaArgTyrGlySerGlyAlaAlaGlyValValAsnIleIleThrLysArgProThr 173
QY 484 GACAAATGGCTCTCTCCGTCATCGAGGCTGAATCTGCAGGAAGCAACAAATGGGT 543
DB 174 AsnAspTrpHisGlySerLeuSerLeuTyrThrAsnGlnProGluSerSerGluGluGly 193
QY 544 AACAGCAGCAGCTTTAAATTTCTGGACGATGTCCTCTGTGGATCAATCTCTCAGCCTG 603
DB 194 AlaThrArgArgAlaAsnPheserLeuSerGlyProLeuAlaGlyAspAlaLeuThrThr 213
QY 604 CAGGTACCGGCTAGCACACACAGCGTCAGGGTTATCGGTACATCATCAGTCAGCGATACA 663
DB 214 ArgLeuTyrGlyAsnLeuAsnLysThrAspAlaAspSer---TrpAspIleAsnSerPro 232
QY 664 GCAGGCACGCGTATTCCTTATCCACGAGTCA---CAGAAATATAATCTTGGTCACGT 720
DB 233 ValGlyThrLysAsnAlaAlaGlyHisGluGlyValArgAsnLysAspIleAsnGlyVal 252
QY 721 CTGTACTGAAAGCGCTGGAGCAGGATGCTCTGCTGTTGATATGATACACCCCGCAG 780
DB 253 ValSerTrpLysLeuAsnProGlnGlnIleLeuAspPheGluValGlyTyrSerArgGln 272
QY 781 -----CGTTATGATAACCGGATGGCACTGGGGAGTCTGACG----- 819
DB 273 GlyAsnIleTyrAlaGlyAspThrGlnAsnSerSerSerSerAlaValThrGluSerLeu 292
QY 820 -----GGGGGATATGACCGGACCTCGCTGCTATGACGGAACAAATTTTCAGCTGCTAT 873
DB 293 AlaLysSerGlyLysGluThrAsnArgLeuTyrArgGlnAsn-----TyrGlyIle 309
QY 874 GATCATATTTTACCTTCGGAACATCGAAATCGTATCTGCACTGGAACAGACAGAA--- 930
DB 310 ThrHisAsn-----GlyIleTrp-----AspTrpGlyGlnSerArgPhe 322
QY 931 -----AAT 933
DB 323 GlyValTyrTyrGluLysThrAsnAsnThrArgMetAsnGluGlyLeuSerGlyGly 342
QY 934 AAGAGTGTGAGTGTGACGATGCTAGTGAACGCGACAAA-----TCGGGGCTT 984
DB 343 GluGlyArgIleLeuAlaGlyGluLysPheThrThrAsnArgLeuSerSerTrpArgThr 362

Alignment Scores:		6.8e-41	Length:	620	
Pred. No.:	Score:	539.50	Matches:	187	
Percent Similarity:		44.69%	Conservative:	141	
Best Local Similarity:		25.48%	Mismatches:	243	
Query Match:		14.46%	Indels:	163	
DB:		6	Gaps:	31	
US-10-625-972-4 (1-2091) x ABM68636 (1-620)					
Qy	7	ATAACCACTCGGCTCCGAGTACCTCCCTGCTCGGATTTC	GCACACAGATA	63	
Db	9	LeuSerThrValSerIleMetValIleSer-----GlyTrpAsnGlnAlaSerAlaA	26		
Qy	64	GCTCCTCGACAGGATGATGATGCTCGGCATCCGGCTATGAGAAAGAGCTGACTAAC	123		
Db	27	ValGluSerGlnAspSerLeuValValThrAlaSerArgPheIleGlnProIleSerSer	46		
Qy	124	GCAGCCCGCAGTCTTCTGTGATTAGCAGAGGAATTGCAGTCCAGCCAGTACCAGAT	183		
Db	47	IleLeuAlaProTyrThrValValThrArgAspGluIleAspArgTrpGlnSerAsnSer	66		
Qy	184	CTGGCGAGGCTCTGAGATCAGTAGAGGCTGCTGATGTT-----GAAAGTGGTACGGGT	237		
Db	67	ValAlaAspIleLeuArgLeuProGlyValAspIleAlaArgHisGlyGlyIleGly	86		
Qy	238	AAAAACCGAGGCTGGAAATCAGCATCCGAGGAATGCCAGCAGTTACACGCTGATCTG	297		
Db	87	GlnLeuSerSerLeu-----PheIleArgGlyThrHisAlaSerHisValLeuValLeu	104		
Qy	298	ATTGATGCTGTTGCTCAG-----GGCGAAGCAGTGCAGTCCCACTCCCAAC	342		
Db	105	MetAspGlyIleArgLeuAsnGlnAlaGlyIleSerGlySerSerAspLeuSer-----	122		
Qy	343	GGTITTTCTGCCATCAATACCGGGTTCATGCCCTCTGGCGGCATTGAGCGATTGAG	402		
Db	123	-----GlnIleProValSerLeuValGlnIleGlyIleGlu	133		
Qy	403	GTTATCAGGGGCGGATCCACACTGTATGGCTCTGATGCGATGGCGGTGCTGGTGAAT	462		
Db	134	TyrIleArgGlyProArgSerAlaValTyrGlySerAspAlaIleGlyIleValIleAsn	153		
Qy	463	ATCATACAGAAAGACACAAATGGCTCTCTCCGTCATGCGAGGCTGAACTG	522		
Db	154	IleIleThrThrArg-----GlnIleLeuGlyThrSerLeuAsnValGlyIleGlySer	171		
Qy	523	CAGGAAAGCAACAAATGGGTAAACAGCAGCCAGGTTTAATTTCTGAGCAGTGGTCCOCTT	582		
Db	172	HisGlyTyrGlnThrTyrAspGlyAlaThrGln-----GlnThrLeu	185		
Qy	583	GTGGATGATCTGTGATGCTCAGCTGAGGTACGGGTACACACACACCGTCAAGTTCATCG	642		
Db	186	AlaGluAsnThrVal---LeuThrAlaAlaAsnTyrThrTyrLysGlyTyrAsp	204		
Qy	643	GTCACATCACTGAGGATACACAGCAGCGGTATTCCTTATCCACCGAGTCAAGAAAT	702		
Db	205	ValValAlaAspGlyAsnThrGlyGlyPheArgGlnProAspArgAspGlyPheMetSer	224		
Qy	703	TATAATCTTGGTGCAGCTCTGACTCGAAGCGGTCCGAGCAG-----	744		
Db	225	LysMetLeuTripleuGlyValAspGlnIlePheAsnGluGlnValSerGlyPheValArg	244		
Qy	745	-----GATGCTCTGG-----	759		
Db	245	AlaTyrGlyTyrAsnAsnArgThrSerTyrAspAlaAspIleAsnTrpSerTyrProTyr	264		
Qy	760	GATATGATACCAACCGCGGCTTATGATATACCGGGATGGCAACTGGGGAGTCTGACG	819		
Db	265	AlaArgProAspThrArgGluLeuTyrSerArgHis-----	276		
Qy	820	GGGGGATATGACCGGACCTCGCTATGACGA-----AACAAATTTTCAGCT	867		
Db	277	-----TyrAspMetGlyValArgPheAsnGlnGlyIleTyrSerSerGlnLeuIleThr	294		
Qy	868	GGCTATGATCATACTTTC-----ACCTTCGGAACATGGAATTCGTAT	909		
Db	295	SerTyrSerHisThrLysAspTyrAsnPheAspProGlnTyrGlyArgTyrAspLysSer	314		
Qy	910	CTGAACCTGAACGACACAGAAAATAAGCTCGTACGCTTGTACCGCAGTGTACTGAAGCGC	969		
Db	315	AlaSerLeuAsnAspSerGluGlnTyrAsnLeuGln-----	326		
Qy	970	GACAAATGGGGCTTCCGGCTCAGCGCGGAGCTTAAGGAATCGAACCTTATCCTGAAT	1029		
Db	327	-----TrpGly-----	329		
Qy	1030	TCATTACTGCTTACCCCTCTCGGAGAAATCTCATCTGTTACGGTGGGGCGGAGTTCAG	1089		
Db	330	ThrPheGlnLeu-----TyrGlnGlyIleValSerThrGlyValAspPheGln	345		
Qy	1090	AGCTCGTCCATCAAGACGAGCTTCTCTTCCGACACAGGTAACACTTCCGGCAGAAA	1149		
Db	346	LysGlnSerIleGluAlaGlyThrSerTyrIleProLysSerLysThrValArgAsnThr	365		
Qy	1150	AGCTGCTCGTATTCTGCTGAGGATGAGTGGCATCTCAGGATGCTACCTCGCTGACTCGG	1209		
Db	366	GlyMetTyrLeuThrAlaGlnGln-----LeuLysAspPheIle-----LeuGluGly	382		
Qy	1210	GGCAGCGCTATGAACATCATGCAATTCGGGGGACACTTCAGTCCGCTGCATATCTG	1269		
Db	383	AlaIleArgSerAspLysHisSerGluAlaGlyTrpAsnThrThrTrpGlnAlaSerLeu	402		
Qy	1270	GTCTGGATGTGGCAGATCCCTGGACGCTGAAAGCGGTGTGACCAACCGGATATAGGCA	1329		
Db	403	GlyTrpGluPheIleLysAspTyrArgLeuIleAlaSerTyrGlyThrAlaPheLysAla	422		
Qy	1330	CCAGAAATGGGCGCTACATAAAGGATAGTGGTGTCTCCGGCAGGGAACAAAT	1389		
Db	423	ProThrLeuSerGlnMetTyr-----GlyPheGly-----	432		
Qy	1390	CTACTTGTAAACCCACCTGAAAGCGGAGAGAGCGCTCAGTTATCAGGCTGGGGTAT	1449		
Db	433	-----GlyAsnHisAspLeuLysProGluGluSerLysGlnTrpGluGlyGlyIle---	449		
Qy	1450	TACGATAACCCCGCTGCAATGCCAATGTCACAGGTTTATGACTGACTCTCTCCAC	1509		
Db	450	-----GluGlyValThrGlyGlnLeuThrTrpArgMetThrValTyrAsnAsn	455		
Qy	1510	AGATTGTCTCTTATTCATTAATGATAACACCAATAGC---TATGTAACACAGCGAAAG	1566		
Db	466	GluIle-----GluGlnLeuIleAspTyrAlaAsnSerArgTyrTyrAsnIleGlyLys	483		
Qy	1567	GCCCGTTCACCGTGTGGAATTTGCCGCACATTCGCCGTGTGTCAGAGGATGTCACG	1626		
Db	484	AlaLysIleLysGlyValGluTrpThrGlyLeuIleAspThrGlyMetPheGlnHisGln	503		
Qy	1627	CTGTCACTGAATTCACCTGGACCCGAGTGAACACCGTATCGTGTGATGATCAACAAAGTGC	1686		
Db	504	LeuThrIleGlnTyrIleAspProArgAsnSerGluThrAsn-----Glu	518		
Qy	1687	CCGTGAGTTATACCCCTGAACACATGGTGAATGCGAAACGAACTGCGAGATCACCGAA	1746		
Db	519	IleLeuValArgAlaLysGlnGlnValLysTyrGlnLeuAspTrpGlnLeuTyrAsp	538		
Qy	1747	GAGTGGCATCATGCTGGTGGTCCCGTTATCCGGGGGAAAACACACCGTTTCCACCAAT	1806		
Db	539	-----PheAspTrp---GlyLeuThrTyrGlnTyrLeuGlyArgArgTyrAspLysAsp	555		
Qy	1807	TATTCTGCTGCTGTACAGAAAGATGTATGATGAGAAAGGAGAAATACCTGAAA	1866		
Db	556	PheSerThrSerProAlaLysArgValLysLeuGlyGly-----ValSer	570		
Qy	1867	GCCTGGACCGTGTGGATGCGAGTCTCTCTGTGGAAGATGACGATGCCCTGACCTGAAT	1926		
Db	571	PheTrp-----AspLeuThrValSerTyrProValThrSerTyrLeuThrIleArg	587		

Db 195 GlnTyrLysTyrArgIleHisLeuAspLysProAsnLeuAsnThrAsnLeuSerLeuGly 214
 Qy 715 GCAGCTTCTGACGAGAGCGTCGGAGGAGTGTCTCGCTTATGATGATACACACC 774
 Db 215 LeuAsnLeuGlyLysPheThrMetAsnThrAspValLeuTyrLysSerPheAsp---Gly 233
 Qy 775 CGGCAGCGTTATGATACCGG----- 795
 Db 234 TyrGlnLeuPheAspLysLysProLeuValLysTyrPheProAlaTyrAsnThrThrIle 253
 Qy 796 -----GATGGGCACTGGGG 810
 Db 254 ThrGluGluLeuSerLysProProThrSerIleSerGlyTyrGluAspValGlnValAla 273
 Qy 811 -----AGTCTGACGGGGGATATGAC 831
 Db 274 HisLysMetAspTyrArgPheSerLysArgLeuLysValGlnLeuLysGlySerTyr--- 292
 Qy 832 CGGACCTGCGCTATGAGCGAAACAAATTCAGCTGGCTATGATCATCTTC----- 885
 Db 293 -----TyrMetLeuAsnLysTyrAspPheGlnAlaAspAsnIlePheGluLys 308
 Qy 886 -----ACCTTCGACATGGAATCGTATCTGAACGGAACGAGACAGAAAT 933
 Db 309 SerGluAspTyrThrTyrGlyGlySerIleAspTyr-----ThrIleSer 323
 Qy 934 AAAGTCTGAGCTTGTACGAGTGTGA-----CTGAAGCGCGCAAAATGG 978
 Db 324 AspLysSerSerLeuValAlaSerValHisThrAspHisTyrAsnArgTyrAspLysTyr 343
 Qy 979 GGGCTT---GCCGTGACCGCGGAGCTTGAAGATCGAACCTTATCTCGAATTCATTA 1035
 Db 344 GluLeuLysSerGlyArgLeuGluTyrLysAsnAsnIleGlnProArgIleVal 363
 Qy 1036 CTGCTTACCCCTCGGAGAAATCTCATCTGTTACGGTGGGGCGGAGTTTCAGAGCTCG 1095
 Db 364 TyrSerThrThrAlaLeuAspLysGlnThrIleThrGlyGlyLeuGluTyrTyrArgGlu 383
 Qy 1096 TCCATGAAGACGGAGTTGCTCTCCAGCAGCAGGTGAACCTTCGGCGCAAAAGCTGG 1155
 Db 384 SerLeuPheSerAspLysPhe-----GluThrGlyValLysGluAsnLysSerGlnTrp 401
 Qy 1156 -----TCGGTATTGTCAGAGTACGAGTGCATCTCAGGATGCATCTGCGCTCACTGCG 1209
 Db 402 TyrAlaThrAlaPheLeuGluAspTrpSerIleAsnLysGlnPheSerValIleAla 421
 Qy 1210 GCGACCCGCTATGAACATCATGACATTCGGGGACATTCAGTCCGCGTGCATATCTG 1269
 Db 422 GlyLeuArgCysAspTyrHisGluLysTyrGlyThrAsnLeuThrProLysAlaSerVal 441
 Qy 1270 GTCTGGGATGGCAGATCCCTGACGCTGAAGCGGTGAAGCGGTGACACGGGATTAAGCA 1329
 Db 442 MetTyrLysIlePhe---ProPheThrValArgPheAsnTyrAlaArgGlyTyrArgSer 460
 Qy 1330 CCCAGATGGCGCAGCTACATAAAGGATTAGTGTGTCTCGGCGAGGAAACAAAT 1389
 Db 461 ProSerIleLysGluLeuTyrMetAsnTrpAspHisLeu-----GlyMetPheTrp 477
 Qy 1390 CPACTTGTAACCCGACCTGAAGCCGGAAGAGAC-----GTCACTTATGAGCTGGG 1443
 Db 478 IleTyrGlyAsnSerLysLysProGluThrAsnAsnTyrIleSerLeu---SerGly 496
 Qy 1444 GTGTATTACGATAACCCCGCTGATGCAATGTCACAGGTTTTCATGACTGACTTC 1503
 Db 497 GluTyrValAsnSerTrpIleAsnIleAsnAlaVal-----TyrSerAsnTrpPhe 514
 Qy 1504 TCCAAACAGATTGCTCTTATTCATAAATGATAACACCAAT---AGCTATGTAACAGC 1560
 Db 515 ArgAsnLysIleGluGlyMetTrpSerAsnAspGlnThrGluLeuHisTyrIleAsnIle 534
 Qy 1561 GGAAGGCCCGGTTCACCGTGTGAATTTGCCGCACATTCGCCCTGTGTCAGAG--- 1617
 Db 535 GlyLysSerArgLeuAlaGlyValGluThrMetCysLysIleGlnIleAsnArgHisIle 554

Qy 1618 GATGTACGCTGTCACTGATTAACACCTGAGCCGGAAGTGAACACGATGGTATTAAC 1677
 Db 555 AsnValHisGlyAlaTyrAsnTyrLeuTyrThr-----SerLysAspAlaAsp--- 570
 Qy 1678 AAAGTGGCGCGCTGAGTTATACCCCTGAACACATGTTGATGCGAACTGAACCTGGCAG 1737
 Db 571 ---GlyValArgLeuSerSerSerProHisSerGlyAsnIleArg----- 585
 Qy 1738 ATCACCAGAGAGTGGCATCATGGTGGTGGCTTATCGCGGAAACACACCGCTTTTC 1797
 Db 586 -----AlaGluTyrAsnThrArgIleProArgTyr 595
 Qy 1798 ACCAG-----AATTATTGCTCACTGAGCGCTGTACAGAGAAA-----GTGTATGAT 1845
 Db 596 AlaThrValValAsnLeuSerGlyAsnIleMetGlyLysLysLysPheAspValLeuAsp 615
 Qy 1846 GAG-----AAAGGAGAAATACCTGAAAGCTCG-----ACGGTGTGTGATGACAGT 1890
 Db 616 GluLeuGluIleAspGlyLysValGluAlaTyrTyrGlnAlaLysValAsnProTyr 635
 Qy 1891 CTGTCTGCGAAGATGACG-----GATGCCCTGACCTGATGATGCTGCG 1932
 Db 636 CysLeuTrpAspLeuThrValSerGlnTyrIleMetGlnAsnLeuArgIleThrAlaGly 655
 Qy 1933 GTGAATAACTGCTCAACAAGGATTACAGTACGAGCGCTGTACAGTCCCGGTAAAGAT 1992
 Db 656 IleThrAsnLeuPheAsp-----TyrThrSerAspArgVal 667
 Qy 1993 AGCTGTATGCGCGTATTACTTCCAGCGGATCATCAACACAGGATATGTGATACCT 2052
 Db 668 Thr-----PheAsnThrSerThrSer-----Pro 675
 Qy 2053 GAGCGAAATTAAGTGTGCTGCTGAACTAT 2082
 Db 676 GlyArgAsnTyrPheIleAlaCysAsnTyr 685

RESULT 9

ID ABU20991 standard; protein; 684 AA.
 XX AC ABU20991;
 XX DT 19-JUN-2003 (first entry)
 XX DE Protein encoded by Prokaryotic essential gene #6518.
 XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
 XX Bacteroides fragilis.
 XX WO200277183-A2.
 XX PN 03-OCT-2002.
 XX PD 21-MAR-2002; 2002WO-US009107.
 XX PF 21-MAR-2001; 2001US-00815242.
 XX PR 05-SEP-2001; 2001US-00948993.
 XX PR 26-OCT-2001; 2001US-0342923P.
 XX PR 08-FEB-2002; 2002US-00072851.
 XX PR 06-MAR-2002; 2002US-0362699P.
 XX PA (ELIT-) ELITRA PHARM INC.
 XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX WPI; 2003-029926/02.
 XX DR N-PSDB; ACA24861.
 XX PT New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to

PT isolate candidate molecules for rational drug discovery programs.

PS Claim 25; SEQ ID NO 48915; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed genetic specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 684 AA;

SQ Alignment Scores:

Pred. No.:	9-52e-28	Length:	684
Score:	399.00	Matches:	164
Percent Similarity:	42.5%	Conservative:	131
Best Local Similarity:	23.6%	Mismatches:	264
Query Match:	10.6%	Indels:	134
DB:	6	Gaps:	34

US-10-625-972-4 (1-2091) x ABU20391 (1-684)

QY	73	GAGGATGTGATG-----ATTGCTCGGCATCCGGCTATGAGAAAAAGCTG	117
DB	72	GluAspGlnMetAsnLeuSerThrValValIleThrGlyThrArgThrProLysLeuLeu	91
QY	118	ACTAACCGGCGCGCCAGTGTTCTGTGATTAGCCAGGAGGAATTGTCAGTCCAGCCAGTAC	177
DB	92	LysAspAlaProIleIleThrArgValIleThrAlaGlyAspLeuLysLysValAspAla	111
QY	178	CACGATCTGGCGGAGGCTCTGAGATCAGTAGAGGGTGTGGANGTTGAAAGTGTACCGGT	237
DB	112	ThrHisIleGlyGlnLeuLeuGln---ValGluLeuProGlyIleGluPheSerTy-Ser	130
QY	238	AAACCGGAGGCTGGAATCAGATCCGAGGAATGCCAGCCATTACACGCTGATACCTG	297
DB	131	MetAspGlnGlnValLysLeuAsnMetGlnGlyPheGlyGlyAsnAlaValLeuPheLeu	150
QY	298	ATTGATGTGTGTCGTCAGGCGGAGCAGTACGTCAGTCCCAACGGTTTCTCGCATC	357
DB	151	ValAspGlyGluArgLeuAlaGlyGluThr---LeuAspAsnIleAspTyAsnArgLeu	169
QY	358	AATACCGGTTTCATGCCCTCTGGCCGATTCAGGGTATTAGGTATCAGGGGGCCG	417
DB	170	Asn-----LeuAspAsnValGluArgValGluIleValLysGlyAla	183
QY	418	ATGTCCACACTGTATGGCTCTGATCGATGGCGGTGTGTGTAATATCATATACCAGAAAG	477

DB	184	AlaSerThrLeuTyGlySerSerAlaIleGlyValIleAsnIleThrLysAla	203
QY	478	AATCCACACAAATGGCTCTCTCCGCAATGATGAGGCTGAATCTGCAGGAAACACACAA	537
DB	204	SerAspAspProTrpAsnLeuAsnLeuAsnThrArgPheGlyValHisAsnAspGlnArg	223
QY	538	TGGGGTAAACAGCAGCCAGTTTAAAT-----TTCTGGAGCAGTGGTCCCTTTGGAT	588
DB	224	HisGlyGlyThrValGlyPheAsnAlaGlyLysPheTy-SerGlnThrAsnValGlnTy	243
QY	589	GATCTCTGC---AGCTGCAGGTACGC---GGTAGCACACACACAGCGTCAGGTTTCATCG	642
DB	244	ThrAsnIleAspSerIleHisValLysGlnGlyAspTyThrThrIleAsnGlyAsnLys	263
QY	643	GTCCATCATCTAGCAGTACATACAGCAGGACCGGCTATTCCTTATCCACGAGTACACAA	702
DB	264	ThrTrpAsnValLysGlu-----ArgLeuMetPhe---ThrProAsnGluGln	278
QY	703	TATAATCTTGGTCAGCTCTTGACTCGAAGCGCTCGGAGCAGGATGTCTCTGTTGAT	762
DB	279	LeuArgLeuThrAlaArgAlaGlyTyTyThrPheArgGluArgAspAla	294
QY	763	ATGGATACACACCGGAGCGGTATGATAACCGGATGGGCAACTGGGGAGTCTGACGGGG	822
DB	295	SerSerGluThrLysAsnArgTy-----ArgGlyPheSerGlyGlyLeuLysGly	311
QY	823	GGATATGACCGGACCTGCGCTATGAGCGAAACAAATTCAGTGGCTATCATCATCT	882
DB	312	AsnTyAsp-----	314
QY	883	TTCACTCTGGACATCGAAATCGTATCTCACTGGAAGACAGACAGACAGAAATAAGTCTG	942
DB	315	-----PheAsnThr---LysSerAsnLeu	321
QY	943	GAGCTTGTACGAGTGTACTGAAGCGCGACAAATGGGG---CTTGCCCGTCAGCGCGG	999
DB	322	GluLeuAlaTyThrPheAspGlnTyAspLysSerAspTyLeuValSerTyLysAsn	341
QY	1000	GAGCTTAAGGATCGAACCTTATCTGGAATTC-----TTACTGCTTACCTCTCTG	1050
DB	342	AspIleArgAspTySerAsnValGlnHisSerVala-GalaLeuTy-AsnTyThrPhe	361
QY	1051	GGAGAATCTCATCTGGTACGGTGGGGCGGAGTTTTCAGAGCTCGTCATGAAAGACGA	1110
DB	362	AsnAspLysAsnThrLeuThrValGlyGlyAspTy-----LeuArgAspTy	377
QY	1111	GTGTCTCTTCCAGCACAGGTGAAACT-----TTCCGCGCAGAAAGCTGTCGGTATTT	1164
DB	378	LeuMetSerTyGlnPheLysGluAsnAlaAspTyThrMetHisSerAlaAspAlaPhe	397
QY	1165	GCTGAGGATGAGTGGCATCTCACGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1224
DB	398	GlyGlnPheAspTrpAsnProThrGluHisPheAsnValIleAlaGlyLeuArgPheAsp	417
QY	1225	CATCATGAGCAATTCGGG---GGACACTCAGTCCGCTGCTGCTGCTGCTGCTGCTGCTG	1281
DB	418	TyrPheSerGluSerAsnValArgHisPheSerProHisLeuGlyLeuMetTyLysIle	437
QY	1282	CGAGATCGCTGCGAGCGTGAAGCGCGGTGTGACACCGGATATAGGCAACCGAGATGGG	1341
DB	438	GlyAsnCys---SerLeuArgGlySerTyAlaGlnGlyPheArgSerProThrLeuLys	456
QY	1342	CAGCTACATAAAGGATAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	1398
DB	457	GluMetHisMetAsnPheTy-----MetAlaAsnThrMetMetIleTyGly	472
QY	1399	RACCCGACCTGACCGGAGAGAGAGCGTCAGTATAG---GCTGGGGTGTATTACGAT	1455
DB	473	AsnProAspLeuGluProGluThrSerHisAsnPheSerLeuSerGlyGlyTyThrLys	492
QY	1456	AAACCCCGCGGCTCTGAATGCCAATGTCAAGGTTTTTATGACTGACTTCTCCACAAAGAT	1515

Db 493 Asn-----ArgTyrAsnPheThrLeuThrGlyTyrTyrAsnLeuValHisAspArgIle 510
Qy 1516 GTCTCTTATTCATTAATGATAAC-----ACCAATAGCTATGTAACACGGGA 1563
Db 511 GlutThrSerPheArgAspThrAspGlyMetIleAlaGlnLysTyrIleAsnThrPro 530
Qy 1564 AAGCCCGGTGACGGTGTGAATTTCCCGGCACATTGCCGTGTGTCAGAGGATGC 1623
Db 531 ArgValAspIleAlaGlyIleAsp---AlaAsnAlaSerAlaLysTyrProCysGlyIle 549
Qy 1624 ACGCTGTCACTGAATTAACCTCGACCGGAAAGTGAACAACGTGATGCTGATTAACAAGGT 1683
Db 550 GlyAlaArgIleSerTyrThrTyrIleHisclupHemMetArgAspGlyGlnThrLys-- 568
Qy 1684 GCGCGCTGAGTTATACCTCGAACAACATGCTGAATCGGAAACTGAACCTGCAGATCAC 1743
Db 569 -----LeuSerSerThrArgProHisSerAlaThrValArgLeuGluTyr----- 583
Qy 1744 GAAGAGTGGCATCATGGCTGGGTGCCCGTTATCGCGGAAACA---CCACGTTTCACC 1800
Db 584 -----GlyLysThrTyrAspHisTyrAsp 591
Qy 1801 CAGAAATTATTCG-----TCACAGAGCGCTGTACAGAAGAAAGTGAT----- 1842
Db 592 PheAsnLeuSerLeuAspGlyArgAlaLeuSerGlnValLysThrAsnGlnTyrThrSer 611
Qy 1843 -----GATGAGAAAGGAAATACCTGAAGCGCTGCAGCGTGGTGGAT 1884
Db 612 AsnAspProAsnAlaGlyThrGluLysValThrTyr---ProGlyTyrThrMetTrpAsn 630
Qy 1885 GCAGGTCTGTGCTGGAAGATGACCGATGCCCTGACGCTGAATGTCGGTGAATAACCTG 1944
Db 631 LeuThrLeuThrGlnArgValTyrPlyGlyIleAsnValAsnMetAlaValAsnLeu 650
Qy 1945 CTCACAAAGGATTACAGTGCAGTGCAGCTGACGCTGACAGTCCGGTAAGAGTACGCTGATGCC 2004
Db 651 PheAsn-----TyrArg 654
Qy 2005 GGTGATTAATCTCCAGACGGGATCATCA---ACAACAGGA 2040
Db 655 ProAspTyrTyrTyrAlaAsnSerProTyrThrThrGly 667
RESULT 10
AY34501
ID AAY34501 standard; protein; 708 AA.
XX AC AAY34501;
XX 27-AUG-2003 (revised)
DT 20-MAR-2003 (revised)
DT 25-AUG-1999 (first entry)
XX Porphorymonas gingivalis protein PG40.
XX Porphorymonas gingivalis; PG; periodontal disease; gingivitis; vaccine;
XX antigenic.
XX Porphorymonas gingivalis.
XX WO929870-A1.
XX 17-JUN-1999.
XX 10-DEC-1998; 98WO-AJ001023.
XX 10-DEC-1997; 97AU-00000839.
PR 31-DEC-1997; 97AU-00001182.
PR 30-JAN-1998; 98AU-00001546.
PR 10-MAR-1998; 98AU-00002264.
PR 09-APR-1998; 98AU-00002911.
PR 23-APR-1998; 98AU-00003128.
PR 05-MAY-1998; 98AU-00003338.
PR 22-MAY-1998; 98AU-00003654.

PR 29-JUL-1998; 98AU-00004917.
PR 30-JUL-1998; 98AU-00004963.
PR 04-AUG-1998; 98AU-00005028.
XX (CSLC-) CSL LTD.
XX Ross BC, Barr IG, Patterson MA, Agius CT, Rothel LJ, Margetts MB;
PI Hocking DM, Webb EA;
XX WPI; 1999-385613/32.
DR N-PSDB; AAX91719.
XX Antigenic Porphorymonas gingivalis peptides for preventing gingivitis.
XX Claim 1; Page 482-483; 588pp; English.
XX AAX91536 to AAX91801 encode two hundred and sixty six antigenic
CC Porphorymonas gingivalis (PG) polypeptide sequences given in AAY34318 to
CC AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
CC activity with a vaccine mechanism of action. The PG polypeptides can be
CC used as vaccines especially against Porphorymonas gingivalis. Probes can
CC be used to detect Porphorymonas gingivalis in standard hybridisation
CC assays. Porphorymonas gingivalis is involved in periodontal disease
CC especially gingivitis. (Updated on 20-MAR-2003 to correct PR field.)
CC (Updated on 27-AUG-2003 to correct OS field.)
XX Sequence 708 AA;
SQ Alignment Scores:
Pred. No.: 1,34e-27 Length: 708
Score: 397.50 Matches: 181
Percent Similarity: 38.88% Conservative: 125
Best Local Similarity: 23.00% Mismatches: 270
Query Match: 10.65% Indels: 211
DB: 2 Gaps: 33
US-10-625-972-4 (1-2091) x AAY34501 (1-708)
Qy 40 CTCGATTTTCAGCCAGCAGCATAGCTGCTGCAGAGGATGTG----- 81
Db 11 IleGlyIleSerAlaGlnAlaPheAlaLysThrAspAsnValProThrAspSerLeuArg 30
Qy 82 -----ATGATTGCTCGGATCGGCTATGAGAAAAGCTGACTAAC 123
Db 31 ValHisAsnLeuGlnThrValThrValTyrSerThrArgThrAlaValProLeuLysLys 50
Qy 124 GCAGCCCGCAGTGTCTGTGATTAGCCAGAGGAAATTCAGTCCAGCCAGTACCAGAT 183
Db 51 IleProAlaLysMetGluLeuIleSerSerArgAsnIleLysGlnSerGlyPheAsnAsn 70
Qy 184 CTGGCGAGGCTCTGAGATCAGTAGAGGCTGGATGTTGAAGTGGTACGGGTAAACC 243
Db 71 MetThrAspIleLeuLysThrGlnSerSerLeuAspValIleGlnTyrProGlyPheSer 90
Qy 244 GGAGGGTGGAAATCAGCATCCCGAGGAATGCCAGCCAGT-----TACACGCTGATACTG 297
Db 91 Ser-----AsnIleGlyIleArgGlyPheLysProSerGlyLysTyrValThrValLeu 108
Qy 298 ATTATGTTGTTCGTTCAGCGCGGAAGCAGTGCAGTCCCAACGGTTTTTTCGCGATG 357
Db 109 ValAsnGlyIleProAlaGly-----ThrAspAsnIleSerThrLeu 122
Qy 358 AATACCGGTTTCATGCCCTCTGGCCCGCATTCAGGCTATTCAGGTTATCAGGGGCGG 417
Db 123 AsnThr-----SerAsnIleGluGlnIleLeuLeuLysGlyPro 136
Qy 418 ATGTCCACACTGTATGCTCTGATGCGATGGCGGTGTGCTGAATATCATACAGAAAG 477
Db 137 PheSerSerIleTyrGlyThrAsnAlaMetGlyValValAsnIleIleThrHisLys 156
Qy 478 AATGCAGACAAA-----TGCGTC 495

Db 157 SerLysAspLysIleHisGlyAsnValSerLeuPheGlySerTyrGlnThrMetAla 176
 QY 496 TCTTCGTCATATCAGG-----CTGAAT 519
 Db 177 GlySerPheAsnLeuGlyArgPheGluAspIlePheSerPheAspLeuSerLeuGly 196
 QY 520 CTCGAGAAAGCAACAAATGGGTACACAGCAGCCAGTTAAATTCCTGGAGCAGTGGTCCC 579
 Db 197 LeuAspLysGlnAsnLysAspTyrLysThrGlySerAsnAsnPheLeuSerLys 216
 QY 580 CTGTGCGATCTCTCTACCTGACGTGACGTACGGGTAGCACAAACAGCGCTCAGGTTCA 639
 Db 217 LeuGluGluAlaIleValAspValAsnAlaThr--LysAsnLysLysMetLysGlySer 235
 QY 640 TCGGTCCACATCACTGACGATACACAGCAGCCGCTATTCTTATCCACAGGAG-- 693
 Db 236 AspTyrThrValAlaThrGlyArgLeuArgPheGlyIleAspPheThrProGluTrpSer 255
 QY 694 -----TCACAGAAATTATACTTGTGTGACACGTCTTGAAGCGGTGCGAGCAG 744
 Db 256 LeuAsnLeuTyrGlnAsnValPheLeuGlyAspAlaIleProValGly----- 271
 QY 745 GATGTCTCTGTTGATATGATACCCCGGACGCTTATGATAACCGGATGGGCAA 804
 Db 271 ----- 271
 QY 805 CTGGGAGTCTGACGGGGGATATGACCGGACCTCGCTATGACGGAACAAATTTCA 864
 Db 272 --GlySerIleTrpGlyValTyrGlyGluSer--LysLysAsnLeuAsnArgSerSer 289
 QY 865 GCTGGCTATGAT-----CATACTTTCACCTTCGGAACATGGAAA 903
 Db 290 ThrSerPheGluLeuLeuGlyLysHisGlyCysHisThrLeuGlnPhe-----Ser 306
 QY 904 TCGTATCTCACTGGAACGACAGACAGAAAT----- 933
 Db 307 ProTyrPheAsnIleGluLysSerGluAsnTyrAsnAsnAlaAspProThrGlyPheIle 326
 QY 934 -----AAAGTCTGTAGCTTGTACGAGTGTACTGAAAGCGACAAATGGGGCTTGCC 987
 Db 327 AsnTyrLysSerAspTyrTyrThrGlyAlaLeuLeuGlnAspLysIleSerPheGly 346
 QY 988 GTTCAGCCCGGAGCTTAAGGAATCGAACCTTATCTCTGAATTCATTACTGTTACCCCT 1047
 Db 347 GlyGln-----AsnIleValLeuGlyVal----- 354
 QY 1048 CTGGGAGATCTCATCTGTTACGTGGGGGCGAG---TTTCAGAGCTCGCTCCATGAAA 1104
 Db 355 -----AspSerArgAsnMetThrMetGluSerGluArgPheGluGlnAla----- 369
 QY 1105 GACGAGTTGCTCTGCGGACGACAGTGAACCTTTCGGCGAGAAAAGCTGCTCGGTATTT 1164
 Db 370 --GlyValAsnThrLysProTyrAsnProGlyTyrAlaThrAsnAsnIleGlyLeuPhe 388
 QY 1165 GCTGAGGATGATGGGAT---CTCAGGATGACACTTGGCTGACTGCGGCGAGCGC--- 1218
 Db 389 GlyGlnAlaAsnPheTyrLeuLeuAsnAspAlaLeuSerIleSerAlaGlyAlaArgAla 408
 QY 1219 -----TATGAACATCATGAGCAA 1236
 Db 409 AspPheMetPhePheAspLeuLysAlaAsnGluTyrLeuAsnAsnGluAlaLysGlnGlu 428
 QY 1237 TTCGGGGACACTTCAGTCCGGGTGCATATCTGCTGGGATGTCGAGATGCTGGAGC 1296
 Db 429 ThrHisAsnValIleAsnProAsnValGlyIleLysTyrGluPheValLysGlyLeuThr 448
 QY 1297 CTGAAAGCGGTGTGACCGGATATAAGGACCCAGATGGGGCGACCTACATAAAGGG 1356
 Db 449 AlaHisGlyThrPheGlySerAlaPheSerAlaPheAspAlaPheGln-----LysAla 466
 QY 1357 ATTAGTGTGTCTCGCGGAGGAAAACAAATCTACTTGTGTAAACCGGACCTGAAGCGG 1416
 Db 467 GlyGlnTyrValGlyProPheGlyThrThr-----IleGlyAsnProAspLeuSerPro 484

QY 1417 GAACAGAGCCTCAGTTATGAGGCTGGGTCTATTACGATAACCC---GCCGCTCGAAT 1473
 Db 485 GluLysSerMetThrTrpAspPheGlyIleGlyTyrSerAsnAlaArgCysGlyIleGln 504
 QY 1474 GCCAATGTC-----ACAGGTTTTATGACTGACTTCTCAACCAAGATTGCTCT--- 1521
 Db 505 AlaAspValThrLeuThrTyrPheHisThrAspHisLysAspLeuLeuSerSerPro 524
 QY 1522 ---TATTCATAAATGATACACCATAGTATGTATAACAGCGGAAGCCCGTTGTCAC 1578
 Db 525 AspTyrAla-----AsnAsnIleThrThrTyrIleAsnAlaAspLysAlaArgMetSer 542
 QY 1579 GGTGTGAA-----TTTCCCGGCACATTGCCCGCTG 1608
 Db 543 GlyIleGluAlaLeuLeuSerTyrAspPheGlySerLeuPheAlaAsnLysPheSerLeu 562
 QY 1609 TGTGCA---GAGGATGTCAGCTGTCTCACTCAATTAACCTGGACCGAAGTGAACAACGT 1665
 Db 563 ArgAlaPheAlaAsnAlaThrIleMetLeuAsnSerGluMetLysLysSerGlnThr--- 581
 QY 1666 GATGGTCAACAAAGGTGCGCGCTGAGTTATACCCCTGAACACATGCTGAATGCGAAA 1725
 Db 582 -----AspAlaProTrpSerGluMetTyrTyrValArgLysGlnAsnIleThrPheGly 599
 QY 1726 CTGAAGTGGCAGATACCGAAGAGTGGCATCATGCTGGTGGTGGCTTATCGCGGAAA 1785
 Db 600 IleGluTyrArgGlyLysGlyGluGluValMetLeuAsnGlyArgPheMetGlyArg 619
 QY 1786 ACACCAAGTTTACCAGAAATTTATTCGTCAGTGGCTGTACAGAGAAAGTGTATGAT 1845
 Db 620 -----ArgIleGluGlnAsnTrp-----TyrAlaTyrTyrProGluValArgPro 634
 QY 1846 GAGAAAGGAGAAATACCTGAAAGCC----- 1869
 Db 635 GluLeuGlnLeuLeuAlaGluGluProGluLeuAlaGlnGlyLeuLeuArg 654
 QY 1870 -----TGGACGTTGGTGGATGAGTCTGTCTGGAAGATGACGATCCCTGAGC 1920
 Db 655 HisProGlnAlaMetValPheAsnAlaSerAlaTyrTyrHisMetAsnLysTyrLeuThr 674
 QY 1921 CTGAATGCTGCGTGAATAACCTGCTCAACAGGATTACAGTACGTCGAGCCTGTACACT 1980
 Db 675 PheGlyValAsnLeuAsnAsnIleLeuAspGluLeuTyrThrGlu----- 689
 QY 1981 GCCGTAAGAGTACGCTGTATGCCGGTGATTCTTCCAGACGGGATCATCAACACAGAA 2040
 Db 690 -----LysaspGly 692
 QY 2041 TATGTGATACCTGAGCGAAT 2061
 Db 693 TyrHisMetProGlyArgAsn 699
 RESULT 11
 ID AAY34374
 XX AAY34374 standard; protein; 772 AA.
 AC AAY34374;
 XX 27-AUG-2003 (revised)
 DT 20-MAR-2003 (revised)
 DT 25-AUG-1999 (first entry)
 DE Porphyromonas gingivalis protein PG40.
 XX Porphyromonas gingivalis; PG; periodontal disease; gingivitis; vaccine;
 KW antigenic.
 XX Porphyromonas gingivalis.
 OS Porphyromonas gingivalis.
 XX WO929870-A1.
 XX 17-JUN-1999.

XX 10-DEC-1998; 98WO-AU001023.
 XX 10-DEC-1997; 97AU-00000839.
 PR 31-DEC-1997; 97AU-00001182.
 PR 30-JAN-1998; 98AU-00001546.
 PR 10-MAR-1998; 98AU-00002264.
 PR 09-APR-1998; 98AU-00002911.
 PR 23-APR-1998; 98AU-00003128.
 PR 05-MAY-1998; 98AU-00003338.
 PR 22-MAY-1998; 98AU-00003654.
 PR 29-JUL-1998; 98AU-00004917.
 PR 30-JUL-1998; 98AU-00004963.
 PR 04-AUG-1998; 98AU-00005028.
 XX (CSLC-) CSL LTD.
 XX Ross BC, Barr IG, Patterson MA, Agius CT, Rothel LJ, Margetts MB,
 PI Hocking DM, Webb EA;
 XX WPI: 1999-385613/32.
 DR N-PSDB; AAX91592.
 XX Antigenic Porphyromonas gingivalis peptides for preventing gingivitis.
 PT Claim 1; Page 338-339; 588pp; English.
 PS AAX91536 to AAX91801 encode two hundred and sixty six antigenic
 CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAX34318 to
 CC AAX34583. AAX91802 to AAX91989 represent PCR primers used in the
 CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
 CC activity with a vaccine mechanism of action. The PG polypeptides can be
 CC used as vaccines especially against Porphyromonas gingivalis. Probes can
 CC be used to detect Porphyromonas gingivalis in standard hybridisation
 CC assays. Porphyromonas gingivalis is involved in periodontal disease
 CC especially gingivitis. (Updated on 20-MAR-2003 to correct PR field.)
 CC (Updated on 27-AUG-2003 to correct OS field.)
 XX Sequence 772 AA;
 SQ

Alignment Scores:
 Pred. No.: 1,418-27 Length: 772
 Score: 397.50 Matches: 181
 Percent Similarity: 38.88% Conservative: 125
 Best Local Similarity: 23.00% Mismatches: 270
 Query Match: 10.65% Indels: 211
 DB: 2 Gaps: 33

US-10-625-972-4 (1-2091) x AAX34374 (1-772)

QY 40 CTCGGATTTCAGCCAGCAGATAGCTGCTGAGAGATGTC----- 81
 DB 75 IIEGLYIIESErAlaGlnAlaPheAlaLysThrAspAsnValProThrAspSerLeuArg 94
 QY 82 -----ATGATTCTCGGCATCCGGCTATGAGAAAAGCTGACTAAC 123
 DB 95 ValHisAsnLeuGlnThrValThrValTyrSerThrArgThrAlaValProLeuLysLys 114
 QY 124 GCAGCGCCGAGCTGTTCTGTGATTAGCAGGAGGAAATTCAGTCAGCCAGTACCAGAT 183
 DB 115 IIEProAlaLysMetGluLeuIleSerSerArgAsnIleLysGlnSerGlyPheAsnAsn 134
 QY 184 CTGGCGGCGCTCTAGATCAGTAGAGGTGTGGATGTGTAAGTGGTACCGGTAAAC 243
 DB 135 MetThrAspIleLeuLysThrGlnSerSerLeuAspValIleGlnTyrProGlyPheSer 154
 QY 244 CGAGGCTGGAATCAGCATCCGAGGATGCCAGCAGT-----TACACGCTGATCTG 297
 DB 155 Ser-----AsnIleGlyIleArgGlyPheLysProSerGlyLysTyrValThrVal 172
 QY 298 ATTGATGTGTTCGTCAGCGCGGAAGCAGTACGTACCTCCCAACGCTTTTCTGCCATG 357
 DB 173 ValAsnGlyIleProAlaGly-----ThrAspAsnIleSerThrLeu 186

QY 358 AATACCGGGTTTCATGCCCTCTGCGCGCCATTGAGCGTATTGAGGTTATCAGGGGGCGG 417
 DB 187 AsnThr-----SerAsnIleGluGlnIleGluIleLeuLysGlyPro 200
 QY 418 ATGTCACACATGTATGGCTCTGATGGATGGCGGTGTGTGAATATCATACCAGAAAG 477
 DB 201 PheSerSerIleTyrGlyThrAsnAlaMetGlyGlyValValAsnIleIleThrHisLys 220
 QY 478 AATGCAGACAAA-----TGGCTC 495
 DB 221 SerLysAspLysIleHisGlyAsnValSerLeuPheGlyGlySerTyrGlnThrMetAla 240
 QY 496 TCTTCGCTCAATGCAGG-----CTGAAT 519
 DB 241 GlySerPheAsnLeuGlyGlyArgPheGluAspIlePheSerPheAspLeuSerLeuGly 260
 QY 520 CTCAGGAAAGCAAAATATGGGTAAACAGCAGCAGCTTAAATTTCTGGAGCAGTGTCTCC 579
 DB 261 LeuAspLysGlnAsnLysAspTyrLysThrGlySerAsnAsnPheLeuSerLys 280
 QY 580 CTTGTGATGATCTCTGTCAGCCTGCAGGTACGCGGTAGCACACACACAGCGTTCAGGTTCA 639
 DB 281 LeuGluGluAlaIleValAspValAsnAlaThr---LysAsnLysLysMetLysGlySer 299
 QY 640 TCGTTCACATCACTGAGCGATACAGCAGCAGCGTATTCTTATATCCACCGAG----- 693
 DB 300 AspTyrThrValAlaThrGlyArgLeuArgPheGlyIleAspPheThrProGluTrpSer 319
 QY 694 -----TCACAGATTATATCTTGTGTCAGCTTTGACTGGAAGCGTGGAGCAG 744
 DB 320 LeuAsnLeuTyrGlnAsnValPheLeuGlyAspAlaIleProValGly----- 335
 QY 745 GATGTGCTCTGTTGATATGATACCCCGCAGCGTTATGATAACCGGATGGGCAA 804
 DB 335 ----- 335
 QY 805 CTGGGAGTCTGACGGGGGATATGACCGACCTCGGTATGAGCGAACAATAATTC 864
 DB 336 ---GlySerIleTrpGlyValTyrGlyGluSer---LysLysAsnLeuAsnArgSer 353
 QY 865 GCTGGCTATGAT-----CATACTTTCACCTTCGGAACATGAAA 903
 DB 354 ThrSerPheGluLeuLeuGlyLysHisGlyCysHisThrLeuGlnPhe-----Ser 370
 QY 904 TCCTATCTGAATCGGAACGACGACAGAAAT----- 933
 DB 371 ProTyrPheAsnIleGluLysSerGluAsnTyrAsnAsnAlaAspProThrGlyPheIle 390
 QY 934 -----AAGGTCTGAGCTTGACGAGTGTACTGAGCGCAGCAAAATGGGGCTTGC 987
 DB 391 AsnTyrLysSerAspTyrTyrThrTyrGlyAlaLeuLeuGlnAspLysIleSerPheGly 410
 QY 988 GGTGAGCGCGGAGCTTAAGGAATCGAACCTTATCTCTGAATTCATTACTGTCTACCCCT 1047
 DB 411 GlyGln-----AsnIleValLeuGlyVal----- 418
 QY 1048 CTGGGAGATCTCATCTGTTAGCTGGGGGGCGAG---TTTCAGAGCTGCTCCATGAAA 1104
 DB 419 -----AspSerArgAsnMetThrMetGluSerGluArgPheGluGlnAla----- 433
 QY 1105 GAGCGAGTCTCTTGTCCAGCAGAGTGAACCTTCGCGCAGAAAGCTGGTCGGTATTT 1164
 DB 434 ---GlyValAsnThrLysProTyrAsnProGlyTyrAlaThrAsnAsnIleGlyLeuPhe 452
 QY 1165 GCTGAGGATGAGTGGCAT---CTCAGGATGCTCGGTGCTGCTGCGGCGCAGCGC--- 1218
 DB 453 GlyGlnAlaAsnPheTyrLeuLeuAsnAspAlaLeuSerIleSerAlaGlyAlaAlaGala 472
 QY 1219 -----TATGAACATCATGAGCAA 1236
 DB 473 AspPheMetPhePheAspLeuLysAlaAsnGluTyrLeuAsnAsnGluAlaLysGlnGlu 492

```

QY 1237 TTGGGGGACACTTCAGTCGGCTGCATATCTGGTCTGGGATGGCAGATCCCTGGACG 1296
Db 493 ThrHisAsnValIleAsnProAsnValGlyIleLysTyrGluPheValLysGlyLeuThr 512
QY 1297 CTGAAAGGGGGTGTGACCCAGGATATAAGGCACCCAGAAATGGGAGCATACATAAAGG 1356
Db 513 AlaHisGlyThrPheGlySerAlaPheSerAlaProAspAlaPheGln-----LysAla 530
QY 1357 ATTAGTGGTGTGTCGGGAGGAGGAAAAAACAATCTACTTGGTAACCCGACCTGAAGCG 1416
Db 531 GlyGlnTyrValGlyProPheGlyThrThr-----IleGlyAsnProAspLeuLysPro 548
QY 1417 GAAGACAGGCTCAGTTATGAGGCTGGGTGTATTACGATAACCC-----GCCGGTCTGAAT 1473
Db 549 GluLysSerMetThrThrPhePheGlyIleGlyTyrSerAsnAlaArgCysGlyIleGln 568
QY 1474 GCAATGTC-----ACAGTTTATGACTGCTCTCCAAACAAGATTGCTCTCT----- 1521
Db 569 AlaAspValThrLeuThrTyrPheHisThrAspHisLysAspLeuIleLeuSerSerPro 588
QY 1522 ---TATTCATAAATGATACACCAATAGCTATGTAACACAGCGGAAAGCCCGGTTGCAC 1578
Db 589 AspTyrAla-----AsnAsnIleThrThrTyrIleAsnAlaAspLysAlaArgMetSer 606
QY 1579 GGTGTGGAA-----TTTCCCGGCACATGCGCGT 1608
Db 607 GlyIleGluAlaLeuLeuSerTyrAspPheGlySerLeuPheAlaAsnLysPheSerLeu 626
QY 1609 TGGTCA--GAGATGTCAGCTGTCTACCTGAATTACCTGGACCCGAGTCAACACGT 1665
Db 627 ArgAlaPheAlaAsnAlaThrIleMetLeuAsnSerGluMetLysLysSerGlnThr--- 645
QY 1666 GATGGTGATAACAAAGGTCGGCTGAGTTATACCTGTAACACATGCTGAATCGAAA 1725
Db 646 -----AspAlaProTrpSerGluMetTyrTyrValArgLysGlnAsnIleThrPheGly 663
QY 1726 CTGAACCTGGCAGATCACCGAAGGTGGCATCATGCTGGTCCCGTATCGCCGGAAA 1785
Db 664 IleGluTyrArgGlyLysGlyLeuGluValMetLeuAsnGlyArgPheMetGlyArg 683
QY 1786 ACACCAAGCTTACCCAGAAATATTCTGCTCACTGAGCGCTGTACAGAAGAAAGTATGAT 1845
Db 684 -----ArgIleGluGlnAsnTrp-----TyrAlaTyrTyrProGluValArgPro 698
QY 1846 GAGAAAGGAGATACCTGAAAGCC----- 1869
Db 699 GluLeuGlnGlnLeuLeuAlaGluGluProGluLeuAlaAlaGlnGlyLeuLeuArg 718
QY 1870 -----TGGACGCTGGTGGATCGAGCTCTGCTGGAAGATCAGGATGCCCTGACG 1920
Db 719 HisProGlnAlaValMetValPheAsnAlaSerAlaTyrTyrHisMetAsnLysTyrLeuThr 738
QY 1921 CTGAATGCTCGGTGAATACCTGCTCAACAAGGATTACAGTACGCTGAGCTGTACAGT 1980
Db 739 PheGlyValAsnLeuAsnIleLeuAspGluLeuTyrThrGlu----- 753
QY 1981 GCCGGTAAGATGACGCTGTATGCCGGTGTATTCTCCAGACGGGATCATCAACAACGA 2040
Db 754 -----LysAspGly 756
QY 2041 TATGTGATACCTGAGCGAAAT 2061
Db 757 TyrHisMetProGlyArgAsn 763

```

RESULT 12

ADC01296

ID ADC01296 standard; protein; 660 AA.

XX AC ADC01296;

XX AC ADC01296;

DT 04-DEC-2003 (first entry)

XX Enterohaemorrhagic E. coli 0157:H7-specific protein SEQ ID NO: 1340.

```

XX enterohaemorrhagic; anti-bacterial.
XX Escherichia coli; 0157:H7.
OS JP2002355074-A.
PN 10-DEC-2002.
PD 24-JAN-2002; 2002JP-00015959.
PF 24-JAN-2001; 2001JP-00112010.
PR (UITS-) UNIV TSUKUBA.
PA WPI; 2003-451640/43.
DR Enterohaemorrhagic Escherichia coli 0157:H7-specific nucleic acid molecule
  and a polypeptide and its use, a polypeptide, a vector and a host cell.
PS Claim 3; SEQ ID NO 1340; 2067pp; Japanese.
XX The invention relates to a novel enterohaemorrhagic Escherichia coli
  0157:H7-specific nucleic acid molecule. A polynucleotide of the invention
  has anti-bacterial activity. The polypeptide can be used in detection
  and/or treatment of 0157:H7 infection. The nucleotide sequence of the
  genome of Enterohaemorrhagic E. coli 0157:H7 was determined. The present
  sequence represents an E. coli 0157:H7-specific polypeptide of the
  invention.
XX SQ Sequence 660 AA;

```

Alignment Scores:

```

Pred. No.: 5,21e-27 Length: 660
Score: 391.00 Matches: 183
Percent Similarity: 39.16% Conservative: 106
Best Local Similarity: 24.80% Mismatches: 257
Query Match: 10.48% Indels: 192
DB: 36

```

US-10-625-972-4 (1-2091) x ADC01296 (1-660)

```

QY 1 ATCGGATAACCACTCTGGCT-----TCGTAAGTCAATCCCTGCTCGGATTTTCA 51
Db 9 LeuArgLeuSerLeuLeuAlaLeuAlaValSerAlaThrLeuProThrPheAlaPhe--- 27
QY 52 GCCAGCAGCATAGCTGCTGCAGAGGATGTCATGATTGCTCGGCATCCGCTATGAGAA 111
Db 28 -----AlaThrGluThrMetThrValThrAlaThrGlyAsnAlaArg 41
QY 112 AAGCTGACTAACCCAGCGCCAGCTGTTCTGTGATTAGCCAGGAGAAATGCGATCCAGC 171
Db 42 SerSerPheGluAlaProMetMetValSerValIleAspThrSerAlaProGluAsnGln 61
QY 172 CAGTACCACCATCTCGCGGAGGCTCTGAGATCAGTAGAGGCTGTGGATGTTGAAAAGTGT 231
Db 62 ThrAlaThrSerAlaThrAspLeuLeuArgHisValProGlyIleThrLeuAsp---Gly 80
QY 232 ACGGGTAAACCCAGGAGGCTGGAATCAGCATCCGAGGAATGCCAGCAGTTACAGCTG 291
Db 81 ThrGlyArgThrAsnGlyGlnAspValAsnMetArgGlyTyrAspHisArgGlyValLeu 100
QY 292 ATACTGATTGATGCTGTCTGTCAGGGCGGAGCAGTGACGTCACTCCCAACGGTTTTTCT 351
Db 101 ValLeuValAspGlyValArgGlnGly-----ThrAspThrGlyHis--- 114
QY 352 GCCATGAATACCGGGTTCATGCCCTCTGCGGCCCATCTGAGCGTATGAGGTATCAGG 411
Db 115 ---LeuAsnGlyThrPheLeuAspPro---AlaLeuIleLysArgValGluIleValArg 132
QY 412 GGGCCGATGTCACACTGCTGATGCGATGGCGGTGTGGTGAATATCATCATTACC 471
Db 133 GlyProSerAlaLeuLeuTyrGlySerGlyAlaLeuGlyGlyValIleSerTyrAspThr 152

```

QY 472 AGAAGAATGAGACAAATGGCTCTCTCCGTCMAATCCAGGGCTGAATCTGCAGGAAGC 531
 Db 153 -----ValAspAlaLysAspLeuLeuGlnGlu--- 161
 QY 532 AACAAATGGGTAAACAGCAGCAGTAAATTTCTGGAGCAGTGGTCCCTTGTGGATGAT 591
 Db 162 -----GlyGlnSerSerGlyPheArgValPheGlyThrGlyGlyThrGlyAspHis 178
 QY 592 TCTGTACGCTG-----CAGGTACGGGTAGCACACACAGCGGTTCAGGTTTCATCGGTC 645
 Db 179 SerLeuGlyLeuGlyAlaSerAlaPheGlyArgThrGluAsnLeuAspGlyIleValAla 198
 QY 646 ACATCACTGACGATACAGCAGGACCGCT-----ATTCTTATCCACGCGAG 693
 Db 199 TrpSerSerArgAspArgGlyAspLeuArgGlnSerAsnGlyGluThrAlaProAsnAsp 218
 QY 694 TCACAGAATTATACTTTGGTGCAGCTTTGACTGGAGGCGTGGAGCAGGATGTGCTC 753
 Db 219 GluSerIleAsnAsnMetLeuAlaLysGlyThrTrpGln----- 231
 QY 754 TGGTTTGATATGATACACCGGCGAGCGTATCATACCGGGATGGCACTGGGAGT 813
 Db 232 -----IleAspSerAlaGln-----SerLeuSerGly 240
 QY 814 CTGACGGGGGATATGACCGGACCTGGCTATGAGGAAACAAATTTTCAGCTGGCTAT 873
 Db 241 LeuValArgTyrTyrAsnAsnAspAlaArgGluProLysAsnProGlnThrValGly--- 259
 QY 874 GATCATACTTTCACCTTCGGAACATGGAAATCGTATCTGAACCTGGACGACAGACAAAT 933
 Db 260 -----AlaSerGluSer 263
 QY 934 AAAGCTGCTGAGCTGTAGCAGTGTACTGAAGCGCGACAAA----- 975
 Db 264 SerAsnProMetValAspArgSerThrIleGlnArgAspAlaGlnLeuSerTyrLysLeu 283
 QY 976 -----TGG----- 978
 Db 284 AlaProGlnGlyAsnAspTrpLeuAsnAlaAspAlaLysIleTyrTrpSerGluValArg 303
 QY 979 -----GGGCTTCCCGTTCAGCGCGGAG-----CTTAAAGAAATCG 1014
 Db 304 IleAsnAlaGlnAsnThrGlySerGlyGlyArgGluGlnIleThrLysGlyAla 323
 QY 1015 AACCTTATCTGAATTCATCTGCTTACCTCCCTCTGGGAGAAATCTCATCTGGTACCGTG 1074
 Db 324 ArgLeuGluAsnArgSerThrLeuPheAlaAspSerPheAlaSerHisLeuLeuThrTyr 343
 QY 1075 GGGGGCGAGTTTCAGAGCTGCTCCATGAAGACGGAGTTGCTTCCCGCAGCAGGTGAA 1134
 Db 344 GlyGlyGlyTyrArgGlnGluGlnHisProGlyGly-----AlaThrThrGly--- 360
 QY 1135 ACTTTCCGCGCAAAAAGC-----TGGTGGTATTGCTGAGCATGAGTGGCATCTC 1185
 Db 361 -----PheProGlnAlaLysIleAspPheSerSerGlyTrpLeuGlnAspGluIleThrLeu 379
 QY 1186 ACGGAT--GCATTTCCGCTGACTCGGCGACCGCTATGACATCATGAGCAATTCGGG 1242
 Db 380 ArgAspLeuProIleThrLeuLeuGlyGlyThrArgTyrAspSerTyrArgGlySerSer 399
 QY 1243 GGACACTTCAGTCCGGTGCATATCTGGTCTGGGATGTG---GCAGATGCCCTGGAGCTG 1299
 Db 400 AspGlyTyrLys-----AspValAspAlaAspLysTrpSerSer 412
 QY 1300 AAAGCGCGTGTGACC-----ACG 1317
 Db 413 ArgAlaGlyMetThrIleAsnProThrAsnTrpLeuMetLeuPheGlySerTyrAlaGln 432
 QY 1318 GGATATAGCCACCCAGATGGGCGAGCTACATAAGGAGTATGCTGTGTCTCC---GGG 1374
 Db 433 AlaPheArgAlaProThrMetGlyGluMetTyrAsnAspSerLysHisPheSerIleGly 452

QY 1375 CAGGGAAGAAACAAATCTA---CTTGTAAACCCCGCAGCTGAAGCCGGAA-----GAGAGC 1425
 Db 453 ArgPheTyrThrAsnTyrTrpValProAsnProAsnLeuArgProGluThrAsnGluThr 472
 QY 1426 GTCAGTTATGAGCGCTGGGTGTATTACGAT-----AACCCCGCGGTCTG 1470
 Db 473 GlnGluTyrGlyPheGlyLeuArgPheAspAspLeuMetLeuSerAsnAspAla---Leu 491
 QY 1471 AATGCCAATGTCCAGGTTTATGACT-----GACTTCTCCAAACAAGATTGTCTCT 1521
 Db 492 GluPheLysAlaSerTyrPheAspThrLysAlaLysAspTyrIleSerThrThrValAsp 511
 QY 1522 TATTCATTAATGATAACCAATAGCTATGTAACAGCGGAAAGCCCGGTGACCGT 1581
 Db 512 PheAlaAlaAlaThrThrMetSerTyrAsnValProAsnAlaLys----- 526
 QY 1582 GTGGAATTTGCCGCGACATTCGCGTGTGGTCAGAGGATGTC----- 1623
 Db 527 -----IleTrpGlyTrpAspValMetThrLysTyrThrThr 538
 QY 1624 ---ACGCTGTCTACTGAATTACACCTGGACCCCGAAGTGAACAACGTCATGGTATACAAA 1680
 Db 539 AspLeuPheSerLeuAspValAlaTyrAsnArgThrArgGlyLysAspThrAspThr--- 557
 QY 1681 GGTGCGCCCTGAGTTATACCCCTGAACACATGCTGAATGCGAACTGAACCTGGCAGATC 1740
 Db 558 GlyGluTyrIleSerSerIleAsnProAspThrValThrSerThrLeuAsnIleProIle 577
 QY 1741 ACCGAAGAGGTGGCATCA-----TGGCTGGGTGGCGGTATATCGCGGGAACACACCAT 1794
 Db 578 AlaHisSerGlyPheSerValGlyTrpValGlyThr---PheAlaAspArgSerThrHis 596
 QY 1795 TTCACCCAGAAATTTCTGTCACCTGACGCTGTACAGACGAAAGTGTATGAGAAAGGA 1854
 Db 597 IleSerSerSerTyrSerLys----- 603
 QY 1855 GATACCTGAAAGCCCTGGACGGTGGTGGATGCAGGTCTGCTGGGAAGATGACGGATGCC 1914
 Db 604 -----GlnProGlyTyrGlyValAsnAspPheTyrValSerTyrGlnGlyGlnAla 621
 QY 1915 CTG-----ACGCTGAATGCTGGGTGAATAACCTGCTCAACAGGATTAC 1959
 Db 622 LeuLysGlyMetThrThrThrLeuValLeuGlyAsnAlaPheAspLysGlyTyr 639

RESULT 13
 ABB52669
 ID ABB52669 standard; protein; 660 AA.
 XX AC ABB52669;
 XX DT 11-FEB-2002 (first entry)
 XX DE Escherichia coli polypeptide SEQ ID NO 732.
 XX KW Escherichia coli; B2/D+A; antiinflammatory; antibacterial;
 XX KW immunosuppressive; extra-intestinal infection; phylogeny; meningitis;
 XX KW systemic infection; non-diarrhoeal infection; septicaemia;
 XX KW pyelonephritis; antibiotic resistance.
 XX OS Escherichia coli.
 XX FN WO200166572-A2.
 XX PD 13-SEP-2001.
 XX PF 12-MAR-2001; 2001WO-EP003445.
 XX PR 10-MAR-2000; 2000FR-00003145.
 XX PR 02-FEB-2001; 2001FR-00001449.
 XX PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 XX PI Bingen E, Bonacorsi S, Clermont O, Nassif X, Tinsley C;

XX WPI; 2001-550253/61.
 XX A library of DNA fragments of *Escherichia coli* strains for the phylogenetic
 PT determination of a given strain comprises polynucleotides of nature B2/D+
 PT A-.
 XX Example 6; Fig 6; 646pp; English.
 XX The invention relates to a library of DNA fragments of *Escherichia coli*
 CC strains comprising polynucleotides (ABA88577-ABA88729 and ABA89533) and
 CC encoded proteins (ABB52459-ABB52919 and ABB52954-ABB53094) of nature
 CC B2/D+A-. The polynucleotides have potential anti-inflammatory,
 CC antibacterial and immunosuppressive activity as part of pharmaceutical
 CC compositions used to treat, palliate or prevent extra-intestinal *E. coli*
 CC infections. The polypeptides are useful for determining the phylogenetic
 CC group of a given *E. coli* strain. These polypeptides can detect and treat
 CC an undesired development of *E. coli*, particularly an extra-intestinal
 CC infection that include systemic and non-diarrhoeal infections such as
 CC septicemia, pyelonephritis and meningitis this is particularly
 CC advantageous as bacterial resistance is increasing with the more frequent
 CC use of broad spectrum antibiotics
 XX Sequence 660 AA;

Alignment Scores:

Pred. No.: 9,94e-27 Length: 660
 Score: 388.00 Matches: 183
 Percent Similarity: 40.17% Conservative: 107
 Best Local Similarity: 25.35% Mismatches: 272
 Query Match: 10.40% Indels: 160
 DB: 4 Gaps: 37

US-10-625-972-4 (1-2091) x ABB52669 (1-660)

QY	1	ATGCGAATAACCACTCTGGCT-----TCGTAAGTCAATTCCTGCTCGGATTTC	51
DB	9	LeuArgLeuSerLeuLeuAlaLeuAlaValSerAlaThrLeuProThrPheAlaPhe---	27
QY	52	GCACGACATAGCTGCTCGACGAGATGATGATGTCGCGCATCCGGCTATGAGAA	111
DB	28	-----AlaThrGluThrMetThrValThrAlaThrGlyAsnAlaArg	41
QY	112	AGCTGACTAAGCAGCCCGAGTCTGCTGATAGCAGGAGGATGAGTCAAGTCA	171
DB	42	SerSerPheGluAlaProMetMetValSerValIleAspThrSerAlaProGluAsnGln	61
QY	172	CAGTACCAGATCTGCGGAGGCTCTGAGATCAGTAGAGGTTGATGTTGAAAGTGGT	231
DB	62	ThrAlaThrSerAlaThrAspLeuLeuArgHisValProGlyIleThrLeuAsp---	80
QY	232	ACGGTAAACCGGAGGCTGGAATCAGCATCCAGGAGATGCCAGCCAGTTACACGCTG	291
DB	81	ThrGlyArgThrAsnGlyGlnAspValAsnMetArgGlyTyrAspHisArgGlyValLeu	100
QY	292	ATACCTGATTGATGTTGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	351
DB	101	ValLeuValAspGlyValArgGlnGly-----ThrAspThrGlyHis---	114
QY	352	GCCATGAATACCGGGTTCATGCCCTCTGCGCGCATTCAGCGTATTGAGGTTATCAGG	411
DB	115	---LeuAsnGlyThrPheLeuAspPro---AlaLeuIleLysArgValGluLeuValArg	132
QY	412	GGGCGGATGTCACACTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT	471
DB	133	GlyProSerAlaLeuLeuTyrGlySerGlyAlaLeuGlyValIleSerTyrAspThr	152
QY	472	AGAAAGATGCAGACAAATGGCTCTCTTCCTGATGAGGAGGAGGAGGAGGAGGAGGAGG	531
DB	153	-----ValAspAlaLysAspLeuLeuGlnGlu---	161
QY	532	AACAAATGGGTAAACAGCAGCAGGTTTAAATTTCTGGACGAGTGGTCCCTTGTGGATGAT	591

DB	162	-----GlyGlnSerSerGlyPheArgValPheGlyThrGlyGlyThrGlyAspHis	178
QY	592	TCTCTACGCTG-----CAGGTACGCGGTAGCACCAACACAGGCTCAGGGTTCATCGGTC	645
DB	179	SerLeuGlyLeuGlyAlaSerAlaPheGlyArgThrGluAsnLeuAspGlyIleValAla	198
QY	646	ACATCACTGAGCATACAGCAGGACGCGT-----ATTCTTATCCACGGAG	693
DB	199	TrpSerSerArgAspArgGlyAspLeuArgGlnSerAsnGlyGluThrAlaProAsnAsp	218
QY	694	TCACAGAAATATATCTTGTGACAGCTCTTGAATGGAAAGCGCTCGGACGAGGAT----	747
DB	219	GluSerIleAsnAsnMetLeuAlaLysGlyThrTrpGlnIleAspSerAlaGlnSerLeu	238
QY	748	-----GTGCTCTGCTTGTATATGATACC-----ACCGG	777
DB	239	SerGlyLeuValArgTyrTyrAsnAsnAspAlaArgGluProLysAsnProGlnThrVal	258
QY	778	CAGCGTTATCATAAACCGGATGGCAACTGGGAGTCTGACGGGGGATATGACGGAC	837
DB	259	GluAlaSerAspSerSerAsnProMetValAspArgSerThrIleGlnArgAlaGln	278
QY	838	CTGCGTATCAGCGAAACAAAATTCAGTGGTGTATGATCATCTTTCACCTTCGGAACA	897
DB	279	LeuSerTyrLys-----LeuAlaProGlnGlyAsnAsp-----	289
QY	898	TGG-----AAATCGTATCTGAACGAAACGACAGCAAAATAAAGTCGTGAGCTT	948
DB	290	TrpLeuAsnAlaAspAlaLysIleTyrTrpSerGluValArg-----	303
QY	949	GTAACGAGTGTACTGAAGCGCGCAAAATGGCGCTTCGCGGTACGCGCGGAG-----	1002
DB	304	-----IleAsnAlaGlnAsnThrGlySerSerGlyGluTyrArgGluGlnIle	319
QY	1003	CTTAAGAAATCGAACCTTATCTGATTCATCTACTCTTACCTTACCTTACCTTACCTT	1062
DB	320	ThrGlyGlyAlaArgLeuGluAsnArgSerThrLeuPheAlaAspSerPheAlaSerHis	339
QY	1063	CTGTTACGCTGGCGGCGGAGTTTCAGAGCTCGTCCATGAAAGACGAGCTTGTCTTGC	1122
DB	340	LeuLeuThrTyrGlyGlyGluTyrArgGlnGluGlnHisProGlyGly-----	357
QY	1123	AGCAGAGTGAACCTTTCCGCGCAAAAGC-----TGTCTGATTTCTGCTGAGGAT	1173
DB	358	ThrThrGly-----PheProGlnAlaLysIleAspPheSerSerGlyTrpLeuGlnAsp	375
QY	1174	GAGTGCATCTACCGAT--GCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1230
DB	376	GluIleThrLeuArgAspLeuProIleThrLeuLeuGlyGlyThrArgTyrAspSerTyr	395
QY	1231	GAGCAATTCGGGACACTTCAGTCCGCTGCATATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1287
DB	396	ArgGlySerSerAspGlyTyrLys-----AspValAspAlaAsp	408
QY	1288	GCTGACGCTGAAAGCGCGTGTGAC-----	1314
DB	409	LysTrpSerSerArgAlaGlyMetThrIleAsnProThrAsnTrpLeuMetLeuPheGly	428
QY	1315	-----ACGGATATAAGGCAACCCAGATGGGCGACCTACATAAAGGATAGTGTGT	1365
DB	429	SerTyrAlaGlnAlaPheArgAlaProThrMetGlyGluMetTyrAsnAspSerLysHis	448
QY	1366	GTGTCC-----GGCAGGAAAAACAATCTA-----CTTGTAAACCGGCTGAGCGGAA	1419
DB	449	PheSerIleGlyArgPheTyrThrAsnTyrTrpValProAsnProAsnLeuArgProGlu	468
QY	1420	-----GAGAGCTCAGTTATGAGCTGGGCTGATTTACGAT-----AAC	1458
DB	469	ThrAsnGluThrGlnGluTyrGlyPheGlyLeuArgPheAspAspLeuMetLeuSerAsn	488
QY	1459	CCGCGCGTCTGAATGCCAATGTACAGGTTTATGACT-----GACTTCTCCAAC	1509
DB	489	AspAla---LeuGluPheLysAlaSerTyrPheAspThrLysAlaLysAspTyrIleSer	507

QY 1510 AAGATTGCTCTTATTCCATAAATGATAACACCAATAGTATGTAAACAGCGGAAAGGCC 1569
 Db 508 ThrThrValAspPheAlaAlaAlaThrThrMetSerTyrAsnValProAsnAlaLys--- 526
 QY 1570 CGGTTGACGGGTGGGAATTGCGGCACATTCGCGTGGTGGTGGAGAGATGTC----- 1623
 Db 527 ---:---:---:---:---:---:---:---:---:---:---:---:---:---:---:---: 534
 QY 1624 -----ACGCTGTCACTGAATTACACCTGGACCGACCGAGTGAACACAGTGAT 1668
 Db 535 LysTyrThrThrAspLeuPheSerLeuAspValAlaTyrAsnArgThrArgGlyLysAsp 554
 QY 1669 GGTGATAACAAAGTGGCGCGCTGAGTTATACCCCTGAAACACATGGTGAATGGGAAACTG 1728
 Db 555 ThrAspThr---GlyGluTyrIleSerSerIleAsnProAspThrValThrSerThrLeu 573
 QY 1729 AACTGGCAGATCACCGAAGGTGGCATCA-----TGGCTGGGTGCCCGTTATCGCGGG 1782
 Db 574 AsnIleProIleAlaHisSerGlyPheSerValGlyTyrValGlyThr---PheAlaAsp 592
 QY 1783 AAAACACACAGCTTTACACCCAGAAATTATTTCGTCACCTGAGCGCTGTACAGAAAGTGAT 1842
 Db 593 ArgSerThrHisIleSerSerTyrSerLys----- 603
 QY 1843 GATGAGAAAGGAGAATACCTGAAAGCCCTGGACGGTGGATGCGAGTCTGTCTGGAAG 1902
 Db 604 -----GlnProGlyTyrGlyValAsnAspPheTyrValSerTyrGln 617
 QY 1903 ATGACGGATGCCCTG-----ACGCTGAATGCTGGGTGAATAACCTGCTCAACAAG 1953
 Db 618 GlyGlnGlnAlaLeuLysGlyMetThrThrThrLeuValLeuGlyAsnAlaPheAspLys 637
 QY 1954 GATTAC 1959
 Db 638 GluTyr 639

RESULT 14
 ADA33859
 ID ADA33859 standard; protein; 643 AA.
 XX
 AC ADA33859;
 DT 20-NOV-2003 (first entry)
 DE Acinetobacter baumannii protein #1020.
 XX
 KW Acinetobacter baumannii; bacterial disease; antibacterial; vaccine;
 KW plant biocontrol agent.
 XX
 OS Acinetobacter baumannii.
 XX
 PN US5652958-B1.
 XX
 PD 13-MAY-2003.
 XX
 PF 04-JUN-1999; 99US-00328352.
 XX
 PR 09-JUN-1998; 98US-0088701P.
 XX
 PA (GENO-) GENOME THERAPEUTICS CORP.
 XX
 PI Breton G, Bush D;
 XX
 DR WPI; 2003-576092/54.
 DR N-PSDB; ADA29733.
 XX
 PT New Acinetobacter baumannii proteins and nucleic acids, useful as reagents
 PT for diagnosing a bacterial disease, as components of antibacterial
 PT vaccines, as targets for antibacterial drugs, or as biocontrol agents for
 PT plants.
 XX
 PS Example; SEQ ID NO 5146; 328pp; English.

XX The invention relates to isolated Acinetobacter baumannii nucleic acids.
 CC The A. baumannii nucleic acids and polypeptides are useful as reagents
 CC for diagnosing a bacterial disease, as components of antibacterial
 CC vaccines, as targets for antibacterial drugs, to detect the presence of
 CC A. baumannii and other Acinetobacter species in a sample, in screening
 CC compounds for the ability to interfere with the A. baumannii life cycle
 CC or to inhibit A. baumannii infection, and as biocontrol agents for
 CC plants. The present sequence represents the amino acid sequence of an A.
 CC baumannii protein.
 XX
 SQ Sequence 643 AA;

Alignment Scores:
 Pred. No.: 2,588-26 Length: 643
 Score: 383.50 Matches: 169
 Percent Similarity: 38.61% Conservative: 109
 Best Local Similarity: 23.47% Mismatches: 279
 Query Match: 10.28% Indels: 163
 DB: 6 Gaps: 32

US-10-625-972-4 (1-2091) x ADA33859 (1-643)
 QY 10 ACCACTCTGGCTTCGCTAGCTCATTCCTGCTCTCGGATTTTCAGCCAGCAGCATACTGCT 69
 Db 31 ThrArgLeuValGlyAlaIleAlaIleAlaMetGly-----CysSerProValIlePhe 48
 QY 70 GCAGAGGATGTG-----ATGATTGCTCGCATCCCGCTAT 105
 Db 49 AlaGluAspAlaThrAspAlaThrGlnLeuAspProIleValIleThrAlaSerLysSer 58
 QY 106 GAGAAATGCTGACTAACCGCAGCGCCAGTGTCTGTGATTAGCCAGGAGAAATTCAG 165
 Db 69 AlaGluLysAlaSerGluValProAlaArgIleSerIleGluProLysIleValGlu 88
 QY 166 TCAGCCAGTACCATCTGCGGAGGCTCTGAGATCAGTAGAGGGTGTGATGTT--- 222
 Db 89 GlnSerProIleAlaGluLeuProHisLeuLeuMetSerAspAlaAlaIleAsnMetVal 108
 QY 223 ---GAAAGTGTACGGGTAAACCGGAGGCTGGAATCAGATCCGAGCAATGCCAGCC 279
 Db 109 GlnSerGlyGlyLeuGlyGlnThrSer-----SerIlePheIleArgGlyThrAsnSer 126
 QY 280 AGTTACACGCTGATCTACTGATTGATGTTCTGTCAGGCGGAGGAGCAGTGCAGTACTCCC 339
 Db 127 GluHisAlaLeuIleLeuArgAspGlyAlaArgLeuAsnThrAlaSerThr----- 143
 QY 340 AACGGTTTTCGCATGATACCGGGTTCATGCCCTCTGCGCCCATTTAGCGCTATT 399
 Db 144 -----GlyAlaAlaAsnLeuAlaPheIle---AspThrThrAspLysIleGlnIle 159
 QY 400 GAGGTTATCAGGGGCGCGATGTCCACACTGTATGGCTCTGATCGATGGCGGTGTGGTG 459
 Db 160 GluIleLeuLysGlyProAlaSerValLeuTyrGlyThrAspAlaIleGlyGlyValVal 179
 QY 460 AATATCATACCAAGAAATGACAGCAATGGCTCTCTCCGTCATCGAGGCTGAAAT 519
 Db 180 GlnIleIleSer---LysThrProGluLysThrSerAlaPheValThrGlyGluIle--- 197
 QY 520 CTGCAGGAACCAACAAATGGGTAAACAGCAGCGACTTTAATTTCTGAGCAGTGTGCC 579
 Db 198 -----GlyGluAsnLysThrTyrLysSerIleValGlyAlaAsp 210
 QY 580 CTTGTGGATGATTTCTTCAGCCTGCGAGGTACGGGTAGCACACACAGCGTTCAGGTTCA 639
 Db 211 LeuAlaGluAsnGlyPheTyrAlaGlnValArgGlyGlnArgLeuGluSerAspGlySer 230
 QY 640 TCGGTACATCACTGAGC-----GATACAGCAGGCGACG 672
 Db 231 ArgIleThrAspLeuLysGlyAsnAspIleLysLysAlaSerTyrAspGlnLysGlyPhe 250
 QY 673 CGTATTCCTTATCCCGAGTACACAGAAATTAATCTTGGTCCACGCTCTTGACTGGAAG 732

251	SerThrIysValGlyValGluLeuAspPhe-----GlyAlaSerValAspTyrThr	268
733	-----CGGTGGAGCAGGATGTGCTCTGGTTGATATGATACACACCGGCAG	780
269	GlnAsnGluGlyThrSerGlnTyrAspThrPheArgTyrAspGlySerLeuThrSerGln	288
781	CGTTATGATAACCGGGATGGCAACTGGGAGTCTGACGGGGGATATGACCGGACCTG	840
289	AspPheLysAsn-----GluLeuLeuAsnIleArgGly-----	299
841	CGCTATGAGCAAAACAAAATTTCAGTGGCTGCTATCATCATCTTTCACTTCCTCGGAACATGG	900
300	-----ArgValAsnLeuAsnSerAspIleSerLeuAsnAlaArgLeuSerGlnPhe	316
901	AAATCGTATCTGAACCTGGACGACAGACAAAATAAAGTCTGTGAGCTTGTACGCAAGTGA	960
317	LysAspGluLeuAsp-----GlnAsnGlySerArgAspPheValHisSerThr	332
961	CTGAAGCGC-----GACAAATGGGGCTTCGGCTCAGCCGCGGAGCTTAAG	1008
333	ThrLysGluAlaGluValTyrGlyIysTrpGlnPheThr-----	345
1009	GAATCGAACCTTATCTCGAATTCACTACTGCTTACCCCTCTGGAGAAATCTCATCTGTT	1068
346	-----SerSerGlnAsnVal	350
1069	ACGGTGGGGCGAGTTTCAGAGCTCGTCCATGAAAGACCGAGTTGTCCTTGCACAGACA	1128
351	LeuAlaGlyValThrHisGlnAsnIle-----AspGlyAspValLeu---SerTyr	366
1129	GGTCAAACTTTCGGCAG-----AAAAGCTGGTGGTATTTCGTAGGATGAGTGGCAT	1182
367	GlySerProTyrAspGluAspValAsnSerThrGlyTyrPheValGlnHisGlnTyrGln	386
1183	CTCAGCGATGCATTCGGCTGACTCGCGGAGCGGCTATGAAACATCATGAGCAATTCGGG	1242
387	---AsnAsnGlyLeuAsnThrGlnValGlyValArgValGluAspAsnGluIysTyrGly	405
1243	CGACATTCAGTCGGCTGCATATCTGGTGGGATGTGCAGATGCTGCAGCTGAA	1302
406	ThrHisThrValAlaGlnGlyAlaIleArgTyrGlnLeuLeuProLeuThrSerIleTyr	425
1303	GGCGGTGTGACCGGATATAGGCACCCAGCAATGGGCGAGTACATAAAGGATTAGT	1362
426	AlaAsnIleGlySerAlaPheLysAlaProThrLeuAsnAspMetTyr-----	441
1363	GGTGTGTCGGGCGAGGAAAAACAATCTACTTGGTAAACCCGACCTGAAGCGGAGAG	1422
442	-----GlySerGly-----GlyAsnProAsnLeuLysProGluGlu	453
1423	AGCGTCAGTTATGAGCTGGGGTG-----TATTACGATAACCCGCGCGTCTG	1470
454	SerIleSerTyrGluValGlyIleAspGlnLysLeuAsnTyrAsnIleSerThrGlyLeu	473
1471	AATGCCAAATCTCACAGTTTATGACTGACTTCTCCACAGAGATTGCTCTTATTCCATA	1530
474	SerAlaTyrTyrThrLysIleAspAsnLeuIleGluSerArgCysIleAlaValCysAsn	493
1531	AATGATAACACCAATAGCTAT-----GTAACAGCGGAAAGCCCGGTTG	1575
494	GlyAspTrpIleAsnThrPheProValTyrGlnAsnIleAsnIleAspLysAlaSerMet	513
1576	CACGGTGTGGAATTCGGCGCACATTGCCGCTGTGTGTCAGAGAT-----GTCACG	1626
514	ArgGlyGlyGluValTyrAlaAsn-----TrpAsnArgAspLeuPheIleLys	530
1627	CTGTCACTGAATPACACCTGGACCCGAGTGAACACAGTGTGGTGATACAAAGGTGCG	1686
531	SerSerTyrAsnTyrValLysAlaIleAsnAspGluThrAspGlnGlu-----	546
1687	CGCGTGAATTATACCCGTGAACACATGGTGAATGCGAAATCGAATCGGAGATCACCGAA	1746
547	---LeuSerArgProArgGlnSerPheThrValSerThrGlyLeuGln-----	562

QY	1747	GAGGTGCATCATGGCTGGGTGCCGCTTATCTCGGGGAAAACACACAGCTTTCACCCAGAAAT	1806	#####
Db	563	-----	-----	AsnGluHis 565
QY	1807	TATTCCG-----	-----	-----
Db	566	TyrGlyLeuSerIleSerLeuSerAla--	-----	-----
QY	1858	TACCTGAAGCCCTGGACGGTGGTGGATGCAGGTCCTGTCGTGGGAAGATGACGGATGCCCTG	1917	-----
Db	585	---ThrProGlyTyrThrThrValaspPheAsnGlyTyrTrpAsnAlaThrProAsnVal	603	-----
QY	1918	ACGCTGAATGCTCGGGTGAATAACCTGCTCAACAGGATTACAGTGACGGTGAGC-----	1971	-----
Db	604	LysLeuPheThrAsnIleGlnAsnIleGlyAspValIlyTyrIlyThrAlaSerTyrAsp	623	-----
QY	1972	-----	-----	-----
Db	624	LysGlyIleTyrTyrValAsnGlyGlyArgLeuAlaSerAlaGlyValThrLeuSerTyr	643	-----
RESULT 15				
ABM67198				
ID	ABM67198	standard; protein; 674 AA.		
XX	ABM67198;			
XX	AC			
XX	AC			
DT	20-NOV-2003	(first entry)		
XX	Photorhabdus luminescens	protein sequence #295.		
XX	Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;			
XX	detection; food; gene expression; plant; animal; macroorganism; toxin;			
KW	antibiotic; biopesticide; virulence factor; disease model; plague;			
KW	whooping cough.			
XX	Photorhabdus luminescens.			
OS	WO200294867-A2.			
EN	28-NOV-2002.			
PD	07-FEB-2002; 2002WO-IB003040.			
PF	07-FEB-2001; 2001FR-00001659.			
XX	(INSP) INST PASTEUR.			
PA	(CNRS) CNRS CENT NAT RECH SCI.			
XX	Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;			
PI	Buchrieser C;			
PI	WPI; 2003-148459/14.			
XX	Genomic sequence of Photorhabdus luminescens and encoded polypeptides,			
XX	useful e.g. as therapeutic antimicrobials and agricultural pesticides.			
XX	Claim 2; SEQ ID NO 295; 1205pp; French.			
XX	The invention relates to the isolation of genes and their encoded			
CC	proteins from Photorhabdus luminescens. The isolated sequences are			
CC	sources of probes and primers for detecting the genome of P. luminescens			
CC	and related species; to study polymorphisms; for gene analysis and for			
CC	detection/amplification of the genes. Antibodies (Ab) raised against the			
CC	polypeptides encoded by the genes are used for detection/identification			
CC	of P. luminescens, e.g. in foods. The genes, proteins, Ab and calls that			
CC	carry a gene-containing vector are used to select compounds that			
CC	modulate, regulate, induce or inhibit expression of the genes in plants,			
CC	animals or microorganisms other than P. luminescens and are able to alter			
CC	response or sensitivity to toxins and antibiotics produced by P.			
CC	luminescens. Cells transformed to express the genes are useful for			
CC	recombinant production of the proteins, particularly toxins and			
CC	antibacterials useful as insecticides, bactericides and fungicides. The			

CC genes, proteins, vectors containing the genes and Ab are also useful
 CC therapeutically (to treat microbial infection by bacteria or fungi that
 CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as
 CC biopesticides. Other uses of the genes and the proteins are as virulence
 CC factors and for identifying targets of human diseases for which P.
 CC luminescens is a model (particularly plague and whooping cough). This
 CC sequence represents one of the isolated P. luminescens proteins
 XX
 SQ Sequence 674 AA;

Alignment Scores:
 Pred. No.: 2,27e-25 Length: 674
 Score: 373.50 Matches: 174
 Percent Similarity: 39.56% Conservative: 131
 Best Local Similarity: 22.57% Mismatches: 273
 Query Match: 10.01% Indels: 193
 DB: 6 Gaps: 36

US-10-625-972-4 (1-2091) x ARM67198 (1-674)

QY 7 ATAACCACTCTGGTTCCTGATGATTCCTGCTCGGATTTTCAGCCAGCAGCATAGCT 66
 Db 19 lIeSerAsnIleProAlaValTyAlaAlaThrLysSerProAspAlaAsnSerAspLys 38
 QY 67 GCTCAGAGGATGATGATGCTCGGCATCCGGCTATGAGAAAGCTACTAACGCA 126
 Db 39 SerSerArgGluValIleThrValTyAlaThrGlyAsnGluArgAspSerPheGluAla 58
 QY 127 GCGCCAGTGTTCGTGATGATGAGGAGGAAATGACGATCCAGCCAGTACCAAGCATG 186
 Db 59 ProMetMetValThrValIleAsnSerGlnSerProGlnSerGlnThrAlaGlyAsnAla 78
 QY 187 GCGGAGGTCTGAGATCAGTAGAGGTGTGATGTTGAAAGTGTGACGGGTAAACCGGA 246
 Db 79 AsnAspLeuIleLysLysIleProGlyIleAsnIle---AlaGlyIleAsnArgAlaAsn 97
 QY 247 GGGCTGGAATCAGCATCCGAGGATGCCAGCAGTACACGCTGATGATGATGATG 306
 Db 98 GlyGlnAspValSerLeuArgGlyTyGlyProLysGlyValLeuThrLeuIleAspGly 117
 QY 307 GTTCTCAGGGGGAAGCAGTGACGTGACTCCCAACGGTTTTTCTGCGCATGAATACCGGG 366
 Db 118 lIeArgGlnGly-----ThrAspThrGlyHis-----IleAsnGlyThr 130
 QY 367 TTCATGCCCCCTCTGCGCCATTCAGCGTATTGAGGTATCAGGGGCGCGATGCCACA 426
 Db 131 PheLeuAspPro---AlaLeuIleLysGlnValGluValAlaValArgGlyProSerAlaLeu 149
 QY 427 CTGTATGCTCTGATGCGTGGCGGTGTGGTGAATATCATCCAGAAAGAAATCCAGAC 486
 Db 150 LeuTyGlySerGlyAlaLeuGlyGlyValIleAlaTyGlnThrValAspAlaAsp 169
 QY 487 AAATGGCTCTCTTCGCTCAATCAGCGGTGAATCTGCAGGAAGCAACAATGGGTAAAC 546
 Db 170 LeuLeuGlnAlaGlyGlnAspHisGlyPheArgValPheAlaArgGlyAlaThrGlyAsp 189
 QY 547 AGCAGCCAGTTTAATTC-----TGGACAGTGTGTCCTTGTGGATGATTC 594
 Db 190 HisSer---MetGlyPheGlyAlaThrPheAlaLysSerGluSerLeuAspGlyVal 208
 QY 595 GTCACTCAGGTACGGGTAGCACACACAGCGTCAAGGTTCATCGGTACATCACTG 654
 Db 209 PheAlaPheSerThrArgAspSerGlyAsnIleArgHisGlyGlyValValGlyAsp 228
 QY 655 ACCGATACAGCAGCGGATTCCTTATCCAGCGAGTCACAAATTAATCTTGGT 714
 Db 229 AsnAspGluPheIleSer-----GCGAGCAGGATGCTC-----AsnLeuMet 237
 QY 715 GCACCTCTTACTGGAAGCGTCGGAGCAGGATGCTC-----ACCGAAGTGAACAA 1662
 Db 238 AlaLysGlySerTrpLysIleAspAsnGlnSerGlnLeuArgTyTy 257
 QY 753 ----- 753

Db 258 ArgAsnGluAlaAsnGlnProLysAsnProGlnThrLeuThrGlyLysAsnLysPro 277
 QY 754 -----TGGTTTCATATGATACCCAGCGAGCGTATGATAACCGGATGGCACTG 807
 Db 278 SerThrTrpThrAspArgThrThrGln-----ArgaspAlaGlnLeu 292
 QY 808 GGGAGTCTCAGCGGGGATATCAGCGACCTCGCTATGAGCGAAACAAAATTTTCAGCT 867
 Db 293 Ser-----TyGlnLeuAsn-----ProAla 299
 QY 868 GGCTATGATCATACTTTACCTTCGGAACATCG-----AAATCGTATCTGACTG 918
 Db 300 GluTyAsp-----TrpLeuAsnAlaLysThrAspLeuTyTy 312
 QY 919 AACGAGACAGAAATAAAGTCTGAGCTTGACGAGTGTACTGAAGCGCAGCAATG 978
 Db 313 SerAspIleThrIleAsnAlaArgThrLysAlaLys----- 324
 QY 979 GGCTTCGCGGTGAGCGGGGAGTTAAGGAATCGAACCTTATCTGATTCATTCATCTG 1038
 Db 325 GlyPheGluGlyArgLysGlnLysThrTyGlyValLysLeuGluAsnArgSerArgLeu 344
 QY 1039 CTTTACCCCTCTGGGAGAATCTCATCTGTTACGTTGGGGGGGAG-----TTTCAGAGCTCG 1095
 Db 345 TrpThrAspSerProLeuAlaHisGlnPheThrTyGlyGlyGluThrTyLysGlnLys 364
 QY 1096 TCCATGAAAGACGAGATGTCCTTGCACAGCAGTGAACCTTCCGGCAG----- 1146
 Db 365 GlnThrProAspGly-----AsnThrAspSerPheProAspAlaAspIle 379
 QY 1147 AAAAGCTGGTGGTATTTCGTGAGGATGAGTGCATCTCCAGGAT---GCACCTGGCGTG 1203
 Db 380 ArgPheTySerGlyTrpLeuGlnAspGluIleThrLeuArgAspLeuProValSerIle 399
 QY 1204 ACTGCGGCGAGCGCTATGAACATCATGACATTCGCGGGACACATTCAGTCGCGTGCA 1263
 Db 400 ValAlaGlyThrArgTyAspAsnTyLys-----GlyAsnAsnSerLysAspGluAla 417
 QY 1264 TATCTGCTGGGATGTCGAGATGCTGAGCCTGAAAGCGGTGTGACC----- 1314
 Db 418 Val-----SerAlaAspLysTrpSerSerLysGlyAlaIleSerIleThrPro 433
 QY 1315 -----ACGGGATATAAGCACCACCGAATGGGG 1341
 Db 434 ThrAspTrpSerMetLeuPheThrSerTyAlaGlnAlaPheArgAlaProThrMetGly 453
 QY 1342 CAGCTATATAAGGATAGTGTGTCGCGGACAGGAAACAAATCTACTTGTGTAC 1401
 Db 454 GluMetTyAsnAspSerMetHisTyProGlyAsnPheTrpThr-----ProAsn 470
 QY 1402 CCGACCTGAAGCGGAGAGCGTCAGTTATGAGCTGGG----- 1443
 Db 471 ProAsnLeuArgProGluSerAsnGluThrTrpGluSerGlyPheGlyLeuArgPheAsn 490
 QY 1444 ---GTGTATTACGATAACCCCGCGTCTGAATGCCAATGTACAGGTTTATGACTGAC 1500
 Db 491 AspLeuAlaAspAsnGlu---LeuLysPheLysAlaSerTyPheAspThrLys 509
 QY 1501 TTCTCCACAGATTTCTCTTATTCCATAAATGATAACACCAAT---ACCTATGTAAC 1557
 Db 510 AlaLysAspTyTrileLysSerAspIleValAsnArgGlyArgAsnThrThrSerValAsn 529
 QY 1558 AGCGAAAGCGCGGTGTCACGCTGGAATTTTCCGCGCAGCATTCGCGTGTGTCAGAG 1617
 Db 530 lIeSerSerAlaLysIleTrpGly-----TrpAspAla 540
 QY 1618 GATGTCAGCTGTCACTGAAT---TACACCTGG-----ACCGAAGTGAACAA 1662
 Db 541 SerIleAsnTyGluSerAsnLeuPheSerTrpGluLeuAlaTyAsnArgThrGluGly 560
 QY 1663 CGTGATGCTGATACAAAGGTGCGCGCTGAGTTATACCTCATCCCTCAACACATGGTGAAT--- 1719

```
Db      561 LysAspGluLysSerGlyGlySerLeuAspAsnLeuSerProAspThrIleThrSerIle 580
QY      1720 -----GCGAAACTGAACTGGCAGATCACCGAAGAGGTGGCATCATGGCTG 1764
Db      581 LeuGluIleProIleAlaGlnThrGlyPheSerVal-----GlyTrpVal 595
QY      1765 GGTGCCCGTTATCGCGGAAACACACCGTTTCACCCAGAAATATTCGTCACCTGAGCGCT 1824
Db      596 Gly-----GlnPheThr---AsnHisThrAspPheLysGly 606
QY      1825 GTACAGAGAAGAGTGTATGATGAGAAAGGAGAAATACCTGAAAGCCTGGACGGTGGTGGAT 1884
Db      607 LysAspAlaArgGlyArgAspProLysGlnGln-----AlaGlyTyrGlyValAsnAsp 624
QY      1885 GCAGGTCTGTCGTGGAG-----ATCAGGATGCCCTGACCGTGAATGCTGCGGTG 1935
Db      625 PheTyrValSerTyrGlnGlyGluGlyMetLeuLysGlyValThrThrAlaValLeu 644
QY      1936 AATAACCTGCTCAACAAGGATTACAGTGACGTGAGCGCTGTACAGTCCCGTAAAGAGTACG 1995
Db      645 GlyAsnAlaPheAspLysGluTyr----- 652
QY      1996 CTGTATGCCGGTGATTACTTCACAGCGGATCATCAACAGGATATGTGATACCTGAG 2055
Db      653 -----TyrSerProGlnGlyThrProGlnAspGly----- 662
QY      2056 CGAAATTACTGGATGTCGCTGAACTATCAGTTC 2088
Db      663 ArgAsnAlaLysLeuPheValSerTyrGlnTrp 673
```

Search completed: October 13, 2004, 11:48:58
Job time : 169 secs

This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:521013
 This read is a RESEQUENCE of a previously sequenced mouse clone
 This read has been verified (found to hit its original self in the correct orientation)
 Seq primer: -4ORP.

FEATURES

source
 Location/Qualifiers
 1. .290
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J x DBA/2J F1"
 /db_xref="taxon:10090"
 /clone="IMAGE:893053"
 /dev_stage="embryo (post-implantation)"
 /lab_host="DH10B"
 /clone_lib="Knowles Solter mouse inner cell mass"
 /notes="Vector: pBluescript SK+; Site1: XbaI; Site2: XhoI; Cloned unidirectionally from mRNA prepared from primitive streak embryonic tissue. Primer: Oligo dt. cDNAs were cloned into the XbaI/XhoI sites of pBluescript SK+ (Stratagene) using commercial linkers (NEB). Average insert size: 0.5 kb."

ORIGIN

Query Match 13.7%; Score 286.8; DB 9; Length 290;
 Best Local Similarity 99.3%; Pred. No. 1.9e-71;
 Matches 288; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 722 TTGACTGGAGGCGTCGGAGCAGATGTCTCTGTTTGTATGATACCCCGCAGC 781
 Db 1 TTGACTGGAGGCGTCGGAGCAGATGTCTCTGTTTGTATGATACCCCGCAGC 60
 QY 782 GTTATGATACCGGGATGGCGAACTGGGAGTCTGACGGGGGATGATACCGACCTTCG 841
 Db 61 GTTATGATACCGGGATGGCGAACTGGGAGTCTGACGGGGGATGATACCGACCTTCG 120
 QY 842 GCTATGACGGAACAAAATTCAGCTGGCTATGATCATCTTTCACCTTCGGAACATGA 901
 Db 121 GCTATGACGGAACAAAATTCAGCTGGCTATGATCATCTTTCACCTTCGGAACATGA 180
 QY 902 AATCGTATCTGACTGGAGCAGACAGAAAATAAGGTCGTGAGCTTGTACGAGTGTAC 961
 Db 181 AATCGTATCTGACTGGAGCAGACAGAAAATAAGGTCGTGAGCTTGTACGAGTGTAC 240
 QY 962 TGAAGCGGCACAAATGGGGGCTTCGGGTCAGCGCGGAGCTTAAGGAA 1011
 Db 241 TGAAGCGGCACAAATGGGGGCTTCGGGTCAGCGCGGAGCTTAAGGAA 290

RESULT 2

AI481927/c
 LOCUS AI481927 289 bp mRNA linear EST 09-MAR-1999
 DEFINITION vH86a12.x1 Knowles Solter mouse embryonic stem cell Mus musculus
 cDNA clone IMAGE:893854 3', mRNA sequence.
 ACCESSION AI481927
 VERSION AI481927.1 GI:4375153
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 289)
 REFERENCE Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pace, D., Harvey, N., Schurk, R., Rittner, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.
 The WashU-NCI Mouse EST Project 1999
 Unpublished (1999)
 CONTACT: Marra M/WashU-NCI Mouse EST Project 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:521814
 This clone was previously sequenced on the 5' end only, this new data is from the 3' end
 Seq primer: Primer name ambiguous.

FEATURES

source

Location/Qualifiers

1. .289
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J x DBA/2J F1"
 /db_xref="taxon:10090"
 /clone="IMAGE:893854"
 /dev_stage="embryo"
 /lab_host="DH10B"
 /clone_lib="Knowles Solter mouse embryonic stem cell"
 /notes="Vector: pSPORT; Site1: NotI; Site2: SalI; Cloned unidirectionally from mRNA prepared from 800 blastocysts. Primer: SalI(dT): 5'-CGTCCACCGTCGACCGTTTTTTTTTTT-3'. cDNAs were cloned into the NotI/SalI sites of a pSPORT vector (Life Technologies)."

ORIGIN

Query Match 13.3%; Score 278; DB 9; Length 289;
 Best Local Similarity 98.3%; Pred. No. 6.7e-69;
 Matches 281; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 775 CGGCAGGTTTATGATACCGGATGGCAACTGGGAGTCTGACGGGGGATATGACCG 834
 Db 289 CGGCAGGTTTATGATACCGGATGGCAACTGGGAGTCTGACGGGGGATATGACCG 230
 QY 835 ACCTGCGCTATGAGCGAAACAAAATTCAGCTGGCTATGATCATCTTACCTTCGGA 894
 Db 229 ACCTGCGCTATGAGCGAAACAAAATTCAGCTGGCTATGATCATCTTACCTTCGGA 170
 QY 895 ACATGGAAATCGTATCTGAACTGGAACGACAGAAAATAAGGTCGTGAGCTTGTACGC 954
 Db 169 ACATGGAAATCGTATCTGAACTGGAACGACAGAAAATAAGGTCGTGAGCTTGTACGC 110
 QY 955 AGTGTACTGAAGCGCGACAAATCGGGGCTTCGCGTCAGCGCGGAGCTTAAGGAATCG 1014
 Db 109 AGTGTACTGAAGCGCGACAAATCGGGGCTTCGCGTCAGCGCGGAGCTTAAGGAATCG 50
 QY 1015 AACCTTATCTGAATTCATCTACTTACCCCTCTCGGAGAACTCTC 1060
 Db 49 AACCTTATCTGAATTCATCTACTTACCCCTCTCGGAGAACTCTC 4

RESULT 3

AZ935601/c
 LOCUS AZ935601 495 bp DNA linear GSS 16-SEP-2002
 DEFINITION T034 E. coli C1845 library subtracted from E. coli MG1655
 Escherichia coli genomic clone T034, genomic survey sequence.
 ACCESSION AZ935601
 VERSION AZ935601.1 GI:14579858
 KEYWORDS GSS.
 SOURCE Escherichia coli
 ORGANISM Escherichia coli
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
 1 (bases 1 to 495)
 REFERENCE Blanc-Potard, A., Tinsley, C., Scaletsky, I., Le Bouguenec, C., Guignot, J., Servin, A., Nassif, X. and Bernet-Camard, M.
 Representational Difference Analysis between Afa/Dr Diffusely Adhering Escherichia coli and Nonpathogenic E. coli K-12
 Infect. Immun. 70 (10), 5503-5511 (2002)
 CONTACT: Blanc-Potard A
 Inserm U510
 Faculte de Pharmacie Paris XI
 92296, Chatenay-Malabry cedex, France

Tel: 33 1 46 83 58 43
Fax: 33 1 46 83 58 44

Email: anne.blanc@cep.u-psud.fr
Class: representational difference analysis.

FEATURES

source

Location/Qualifiers

1. 495
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="C1845"
/db_xref="taxon:562"
/clone="T034"
/clone_lib="E. coli C1845 library subtracted from E. coli MG1655"

ORIGIN

Query Match 12.0%; Score 251; DB 28; Length 495;
Best Local Similarity 90.0%; Pred. No. 6.2e-61;
Matches 280; Conservative 0; Mismatches 30; Indels 1; Gaps 1;
QY 1753 GCATCATGGCTGGTGGCCGTTATCGCGGGAACACACCGTTTCAACCAGAAATTATTCG 1812
DB 311 GCGTGTGTCACCTGGCGGGTTCGACGCGGTACAAACGTCGCCAGCGTGCAATTATTCG 252
QY 1813 TCATGACCGCTGTACAGAAAGATGTATGATGAGAAAGAGATACCTGAAAGCCTGG 1872
DB 251 TCACTGACGCTGTACAGAAAGATGTATGATGAGAAAGAGATACCTGAAAGCCTGG 192
QY 1873 ACGTGTGGATGACGAGTCTGTCTGGGAAGATGACGATGCGCTGAATGCTGGC 1932
DB 191 ACGTGTGGATGACGAGTCTGTCTGGGAAGATGACGATGCGCTGAATGCTGGC 132
QY 1933 GTGAATAACCT-GCTCAACAGGATTACAGTACGATGAGCCTGTACAGTCCGCTAAGAG 1991
DB 131 GTGAATAACCTAGCTCAACAGGATTACAGTACGATGAGCCTGTACAGTCCGCTAAGAG 72
QY 1992 TACGCTGTATCCGCTGATTCTCCACGCGGATCATCAACAGGATGATGATACC 2051
DB 71 TACGCTGTATCCGCTGATTCTCCACGCGGATCATCAACAGGATGATGATACC 12
QY 2052 TGAGCGAAATT 2062
DB 11 TGAGCGAAATT 1

RESULT 4

AA501138

LOCUS

DEFINITION

vh77h07.r1 Knowles Solter mouse inner cell mass Mus musculus cDNA

clone IMAGE:893053 5', mRNA sequence.

ACCESSION

AA501138

VERSION

AA501138.1

KEYWORDS

EST.

SOURCE

Mus musculus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 243)

REFERENCE

AUTHORS

Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,

Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,

Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and

Waterston, R.

The WashU-HMI Mouse EST Project

Unpublished (1996)

Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.lnl.gov) for further information.

FEATURES

source

Location/Qualifiers

1. 243
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J x DBA/2J F1"
/db_xref="taxon:10090"
/clone="IMAGE:893053"
/dev_stage="embryo (post-implantation)"
/lab_host="DH10B"
/clone_lib="Knowles Solter mouse inner cell mass"
/note="Vector: pBluescript SK+; Site 1: XbaI; Site 2: XhoI; Cloned unidirectionally from mRNA prepared from primitive streak embryonic tissue. Primer: Cligo dr. cDNAs were cloned into the XbaI/XhoI sites of pBluescript SK+ (Stratagene) using commercial linkers (NEB). Average insert size: 0.5 kb."

ORIGIN

Query Match 10.7%; Score 224.6; DB 9; Length 243;
Best Local Similarity 97.9%; Pred. No. 1.8e-53;
Matches 238; Conservative 0; Mismatches 4; Indels 1; Gaps 1;
QY 803 AACTGGGGAGTCTGACGGGGGATATGACGGACCTCGGCTATGAGCGAA-ACAAAATT 861
DB 1 AACTGGGGAGTCTGACGGGGGATATGACGGACCTCGGCTATGAGCGAA-ACAAAATT 60
QY 862 TCAGCTGGCTATGATCATATCTTTCACTTCGGAACATGGAATCGTATCTGAACCTGGAAC 921
DB 61 TCAGCTGGCTATGATCATATCTTTCACTTCGGAACATGGAATCGTATCTGAACCTGGAAC 120
QY 922 GACACAGAAATAAGGTGTGAGCTGTACCGAGTGTACGAGCGGACCAAAATGGGG 981
DB 121 GACACAGAAATAAGGTGTGAGCTGTACCGAGTGTACGAGCGGACCAAAATGGGG 180
QY 982 CTTGCCCGCTCAGCGCGGGAGCTTAAAGGAATCAACCTTATCTGAATTCATTACTGCTT 1041
DB 181 CTTGCCCGCTCAGCGCGGGAGCTTAAAGGAATCAACCTTATCTGAATTCATTACTGCTT 240
QY 1042 ACC 1044
DB 241 ACC 243

RESULT 5

AQ990279

LOCUS

DEFINITION

Photobacterium luminescens strain W14 M13 library

Photobacterium luminescens genomic clone pLG01023, genomic survey

sequence.

ACCESSION

AQ990279

VERSION

AQ990279.1

KEYWORDS

GSS.

SOURCE

Photobacterium luminescens

ORGANISM

Photobacterium luminescens

Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

Enterobacteriaceae; Photobacterium.

1 (bases 1 to 586)

REFERENCE

AUTHORS

ffrench-Constant, R.H., Waterfield, N., Burland, V., Perna, N.T.,

Daborn, P.J., Bowen, D. and Blattner, F.R.

A genomic sample sequence of the entomopathogenic bacterium

Photobacterium luminescens W14: potential implications for virulence

Appl. Environ. Microbiol. 66 (8), 3310-3329 (2000)

JOURNAL

MEDLINE

PUBMED

COMMENT

Contact: ffrench-Constant RH

Department of Biology and Biochemistry

University of Bath

South Building, Bath BA2 7AY, UK

Tel: (44) 1225 826621

Fax: (44) 1225 826779

Email: bsr@bath.ac.uk

This is one of 2,122 random reads from the M13 library. For

annotation of identified clones (BLASTX, BLASTN and mapping to E. coli K12 genome) please see french-Constant et al. 2000, Nucleic Acids Res.

Seq primer: M13 Forward
Class: shotgun.

FEATURES
source
1. .586
Location/Qualifiers
/organism="Photorhabdus luminescens"
/mol_type="genomic DNA"
/strain="W14"
/db_xref="taxon:29488"
/clone="PLG01023"
/dev_stage="primary phase variant"
/clone_lib="Photorhabdus luminescens strain W14 M13 library"
/notes="Genomic DNA from strain W14 was size selected (1-2 kb) and then cloned into M13 Janus."

ORIGIN

Query Match 3.9%; Score 81.8; DB 28; Length 586;
Best Local Similarity 53.4%; Pred. No. 6.8e-12; Indels 6; Gaps 1;
Matches 198; Conservative 0; Mismatches 167;
QY 1150 AGCTGCTGGTATTTCTGAGGATGAGTGGCATCTCACGGATGCACCTTTCGCTGACTGCG 1209
Db 133 AGTTGGCGTGTGTTCTGAGGATGAGTGGCATGACCAATGACTTTTCTTCTGACTGGT 192
QY 1210 GGCAGCGGTATGAACATCATGAGCAATTCGGGGACACATTCAGTCCGGTGCATATCTG 1269
Db 193 GGTATCCGTATGATTAAGACGAGAAATTTGGCAGTCATTTGACCTCCACGCTTATATGGT 252
QY 1270 GTCTGGGATGTGGCAGATGCTGGACGCTGAAAGCGGTGACCAACGGATATAAGGCA 1329
Db 253 GTGTGCATGTAGATGAACATGGAATTAAGGTGGTGTATCTACAGTTACCGTTCT 312
QY 1330 CCAGAAATGGGCA-----GCTACATAAAGGATAGTGGTGTGTCCTGGGAGGCAAA 1383
Db 313 CCGGATTTGCTTATATTTTCGGCGGATTCGGGCGAGCAAAACAGGTGTTATGTTGGAAAT 372
QY 1384 ACAATCTACTTGGTAAACCCGACCTGAAGCGGAGAGAGCGTCACTGTTATGAGCTGGG 1443
Db 373 GGTATTATTTCTGGTAACCGAATTTGAACACAGAGAAAGTCAATGAAGAAATTAAT 432
QY 1444 GTGATTACCAATACCCCGCTGCAATGCCAATGTACAGGTTTATGACTGACTTC 1503
Db 433 GTATGTGAACAATCCGGATAATTTAATGCTGGATTGACCATATTTAATACTGATTT 492
QY 1504 TCCAAACAGAT 1514
Db 493 AAGATAAAAT 503

RESULT 6
CF199793/c
LOCUS
DEFINITION
EST1397 Tamarix androssowii leaf Tamarix androssowii CDNA, mRNA
sequence.
ACCESSION
CF199793
VERSION
CF199793.1 GI:33394166
KEYWORDS
EST.
SOURCE
Tamarix androssowii
ORGANISM
Tamarix androssowii

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Caryophyllales; Tamaricaceae; Tamarix.

REFERENCE
1 (bases 1 to 690)
EST acquired from cDNA library of Tamarix androssowii treated with NaHCO3
Unpublished (2003)

JOURNAL
AUTHORS
TITLE
COMMENT
Contact: Yucheng Wang
Forestry Source and Environment College
Northeast Forestry University

Hexing 26, Harbin, Heilongjiang, 150040, P.R. China
Tel: 086-451-2190607
Email: WANGYUCHENG1029@YAHOO.COM.CN.

FEATURES
source
1. .690
Location/Qualifiers
/organism="Tamarix androssowii"
/mol_type="mRNA"
/db_xref="taxon:189785"
/tissue_type="leaf"
/clone_lib="Tamarix androssowii leaf"

ORIGIN

Query Match 3.2%; Score 67.4; DB 14; Length 690;
Best Local Similarity 53.6%; Pred. No. 1.1e-07;
Matches 140; Conservative 0; Mismatches 121; Indels 0; Gaps 0;
QY 388 ATTGAGCGTATTGAGGTATATCAGGGGCGGATGTCACACTGTATGGCTCTGTATGCGATG 447
Db 302 ATTGAACGTATTGAAGTTCTGCGTCCGCGAGCTGCGCTTATGCAACGCGCGCG 243
QY 448 GCGGTGCTGTAATCATTTACAGAAAGATGCAGACAAATGGCTCTCTCCGTCAT 507
Db 242 GCGGCGGTGTTAACTATCATTTACAAAAGGAGCGCGAGTGGCAGCTCTCTGGAC 183
QY 508 GCAGGGCTGAATCTGCAGGAAAGCAAAATGGGGTAACAGCAGCCAGTTTAAATTTCTGG 567
Db 182 GCATATTTCAATCGCCAGAACATAAAGAGGAAGTGCACCAACGCACTTAATTTAGC 123
QY 568 AGCATGGTCCCTTTGGTATGATTTCTGTCAGCTTCAGGTACGCGGTAGCACACACAG 627
Db 122 CTGACCGTCCGTCGGCGACGAATTCGACAGAGCTTGTATCAATTTGTGCTCTATAC 63
QY 628 CGTCAGGGTTTCATCGGTCACA 648
Db 62 TAAACAAGTGTAACTGTACA 42

RESULT 7
BZ576672/c
LOCUS
DEFINITION
msh2_5047.y2 msh Pseudomonas aeruginosa genomic clone msh2_5047,
genomic survey sequence.
ACCESSION
BZ576672
VERSION
BZ576672.1 GI:27211733
KEYWORDS
GSS.
SOURCE
Pseudomonas aeruginosa
ORGANISM
Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.

REFERENCE
1 (bases 1 to 988)
AUTHORS
Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE
Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
J. Bacteriol. (2002) In press
CONTACT: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.

Location/Qualifiers
1. .988
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="WSH"
/db_xref="taxon:287"
/clone="msh2_5047"
/clone_lib="msh"
/notes="Environmental isolate. Whole genomic shotgun
library."

FEATURES
source

```

ORIGIN
Query Match          3.1%; Score 64.8; DB 28; Length 988;
Best Local Similarity 55.9%; Pred. No. 7.9e-07;
Matches 123; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

QY 1199 CGCTGACTGCGGCGAGCGCTATGAAATCATATGAGCAATTCGGGGGACACTTCAGTCCGC 1258
      |||||
Db 392 CGCTGACCTCGGAGAGCGCTTCACACAGCAGAGGCGTTCGGTGGGAAAGCAGTCCGC 333

QY 1259 GTGCATATCTGCTCTGGGATGTGCAGATGCTCGAGCGCTGAAAGGCGGTGTGACACAGG 1318
      |||||
Db 332 GCCTGTACTCTGCTGACACACCTTTCCGATGCGTGAACCTGCGGCGCGGCGTCCGCGCG 273

QY 1319 GATTAAGGACCCAGATGCGGAGCTACATAAGGATAGTGGTGTGTCGGGCGAGG 1378
      |||||
Db 272 GCTCAAGCGCCGAGCCTGAAGCAGCTATCCGCGGAGTATGCGGCGTGGTGGCGCG 213

QY 1379 GAAAAACAAATCTACTTGGTAACCCCGACCTGAAAGCCGGA 1418
      |||||
Db 212 GCGCTTCCCATCTACGCGCAATCCGACCTGAAAGCCGGA 173

RESULT 8
CG745412/c
LOCUS
DEFINITION
P038-2-C09.za Ppa EcoRI BAC library Pristionchus pacificus genomic,
Genomic survey sequence.
ACCESSION
CG745412
VERSION
CG745412.1 GI:37966338
KEYWORDS
GSS.
SOURCE
Pristionchus pacificus
ORGANISM
Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
Neodiplogasteridae; Pristionchus.
REFERENCE
1 (bases 1 to 1350)
Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K.,
Buntjer,J., van der Meulen,M. and Sommer,R.J.
An integrated physical and genetic map of the nematode Pristionchus
pacificus
Mol. Genet. Genomics 269 (5), 715-722 (2003)
22835951
12884007
Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spenannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: raif.sommer@tuebingen.mpg.de
Class: BAC ends.
FEATURES
Location/Qualifiers
source
1..1350
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/clone_lib="Ppa EcoRI BAC Library"
/note="The library was generated by a partial digest of
the genomic DNA with EcoRI and cloning into the BAC
vector."

ORIGIN
Query Match          3.0%; Score 62.4; DB 29; Length 1350;
Best Local Similarity 57.0%; Pred. No. 4.7e-06;
Matches 114; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 388 ATTGAGCGTATTGAGTTATCAGGGCGGATGTCCACACTGTATGCTCTGATCGGATG 447
      |||||
Db 205 ATTGACGATTGAGTTCTGCTGGTCCGGAGCTGCGGTTATGGCAAGCGCGCGG 146

QY 448 GCGGCTGCTGTAATATATATACCAAGAAAGATGCAGACAAATGCTCTCTTCCTCAAT 507
      |||||

```

```

Db 145 GGCGGCGTGGTTAAATCATATTACCAAAAAGGACGGCGAGTGGCAGCGCTCTCTGGGAC 86
QY 508 GCAGGCGCTGAATCTGCAGGARRAGCAACAAATCGGTAAACAGCAGCCAGTTTAAATTTCTGG 567
      |||||
Db 85 GCATATTTCAATGGCGGAGACATTAAGAGGAGGTGCCACCAACGCACTTAATTTAGC 26
      |||||
QY 568 AGCAGTGTGTCCTCTTTGTGGA 587
      |||||
Db 25 CTGACCGGTCCGCTGGGCGA 6

RESULT 9
AZ302660/c
LOCUS
DEFINITION
GSSBrui597 Brucella abortus random genomic library Brucella
melitensis biovar Abortus genomic clone UU1597, genomic survey
sequence.
ACCESSION
AZ302660
VERSION
AZ302660.1 GI:10128871
KEYWORDS
GSS.
SOURCE
Brucella melitensis biovar Abortus (Brucella abortus)
ORGANISM
Brucella melitensis biovar Abortus
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Brucellaceae; Brucella.
REFERENCE
1 (bases 1 to 519)
Sanchez,D.O., Zandomeni,R.O., Cravero,S., Verdun,R.E., Pierrou,E.,
Faccio,P., Diaz,G., Lanzavecchia,S., Aguero,F., Frasch,A.C.C.,
Andersson,S.G.E., Rosetti,O.L., Grau,O. and Ugalde,R.A.
Gene discovery through genomic sequencing of Brucella abortus
Infect. Immun. 69 (2), 865-868 (2001)
21101034
11159979
Contact: Siv Andersson
Small Genomes Sequencing Group
Department of Molecular Evolution, Uppsala University
Norbyvagen 18C, S-752 36, Uppsala, Sweden
Tel: 46-18-471-4379
Fax: 46-18-471-6404
Email: Siv.Andersson@ebc.uu.se
Sequences were basecalled with phred and vector was masked with
crossmatch (see http://genome.washington.edu). Sequences were then
trimmed from both ends to remove low quality bases and masked
vector.
Class: shotgun.
FEATURES
Location/Qualifiers
source
1..519
/organism="Brucella melitensis biovar Abortus"
/mol_type="genomic DNA"
/strain="2308"
/db_xref="taxon:235"
/clone_lib="UU1597"
/clone="Vector: modified M13"

ORIGIN
Query Match          2.7%; Score 57.2; DB 28; Length 519;
Best Local Similarity 72.5%; Pred. No. 8.4e-05;
Matches 74; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 376 CCTCTGCGCCCGCAATGAGCGTATGAGTTATCAGGGGCGGATGTCACACTGTATGGC 435
      |||||
Db 354 CGCTGACCTCCCATCGAACGTTATCGAGATTGCGGGGCGGCGATTCTCTCGCAATATGGC 295

QY 436 TCTGATCGATGCGGCGGTGGTGAATATCATCATTACCAGAAAG 477
      |||||
Db 294 GCGGATCGAATCGGCGGTGTCATCAATATCATCACCACAG 253

RESULT 10
BZ568287
LOCUS
DEFINITION
BZ568287 1029 bp DNA linear GSS 17-DEC-2002
pacs2-164_7457.y2 pacs2-164 Pseudomonas aeruginosa genomic clone
pacs2-164_7457, genomic survey sequence.

```

ACCESSION BZ568287
VERSION BZ568287.1 GI:27201012
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.

REFERENCE 1 (bases 1 to 1029)
AUTHORS Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M., Burns, J.L., Kaul, R. and Olsen, M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of Pseudomonas aeruginosa library
JOURNAL J. Bacteriol. (2002) In press
COMMENT Contact: Chris K. Raymond
 Genome Center
 University of Washington
 Box 352145, Seattle, WA 98105-2145, USA
 Tel: 2062216954
 Fax: 2066857244
 Email: craymond@u.washington.edu
 Class: shotgun.

FEATURES
 source
 1..1029
 /organism="Pseudomonas aeruginosa"
 /mol_type="genomic DNA"
 /strain="2-164"
 /db_xref="taxon:287"
 /clone="pacs2-164_7457"
 /clone_lib="pacs2-164"
 /note="clinical isolate 2-164 Whole genomic shotgun library."

ORIGIN
 Query Match 2.6%; Score 55.2; DB 28; Length 1029;
 Best Local Similarity 52.3%; Pred. No. 0.00049;
 Matches 201; Conservative 0; Mismatches 168; Indels 15; Gaps 3;

QY 75 GGATGTGATGATGTCGCGCATCGCGCTATGAGAAAAGCTGACTAACGCGAGCGCCGAG 134
 DB 264 GGAGCGGACCGTGTGTCAGCGCCACCACTACCGAGCGCAAGCTGCGCGAGCGCCGCGGAG 323
 QY 135 TGTTCCTGTGATAGCAGGAGAAATTCAGTCCAGCCAGTACCACTGTCGGGGAGGC 194
 DB 324 CGTTTCGCTGATCGCGCGGAGGACCTGCGCGGCGCGCGCTGCGCGGACCTGCGAGGAGC 383
 QY 195 TCTGAGATCAGTAGAGGGTGTGGAATGTTGAAAAGTGTACGGGTAAACCGGAGGCGCTGGA 254
 DB 384 CTTGCGCGGTTTCGGAAGCCTGAGTTCAACGGCTCGGC-----ATGTCGCGCGGG 437
 QY 255 AATCAGCATCCGAGGATCCAGCCAGTTACACGCTGATATGATGATGATGTTTCGTCA 314
 DB 438 CATCAGCGTTCGCGGATCTCCAGCGAGCACACCTGCTGTTGTTGGAGCGGC-----CA 491
 QY 315 GGGCGGAAGCAGTACGCTGACTCCCAACGGTTTCTGCCATGAATACCGGTTTCATGCC 374
 DB 492 GGGATCTCCACTCGTTCGGGGCCATCGCCCATTCGCACTTCGACCTCGGCTGGTG-- 549
 QY 375 CCCTTCGCGCGCATGAGCGTATGAGGTTATCAGGGGCGGATGTCACACTGATG 434
 DB 550 -CCGGTGGAGAGCATCGAGCGTATCGAGGTGTACGGGACCGATGCTCGCTCTACGG 608
 QY 435 CTCTGATCGATGGCGGGTGTGTT 458
 DB 609 TTCGGAAGCCCTCGCGGGGTGT 632

RESULT 11
CG891863/c
LOCUS CG891863
DEFINITION 01S0518-03C1-H08 UniformMu MutAIL Library Zea mays genomic clone
ACCESSION CG891863
VERSION CG891863.1 GI:38646060

KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 351)
AUTHORS Latschew, S., Tan, B.-C., Settles, A.M. and McCarty, D.R.
TITLE Sequence tagged transposon insertions from the UniformMu maize population
JOURNAL Unpublished (2003)
COMMENT Contact: Donald R. McCarty
 Plant Molecular and Cellular Biology Program
 University of Florida
 PO 110690 Gainesville, FL 32611-0690, USA
 Tel: 352-392-1928 x322
 Email: drmc@ufl.edu
 Sequence flanking probable Mu insertion site in UniformMu line: 01S0518-03, Primer set: C
 Class: transposon insertion site.

FEATURES
 Location/Qualifiers
 1..351
 /organism="Zea mays"
 /mol_type="genomic DNA"
 /strain="W22 (ACR, bz1-m9)"
 /cultivar="UniformMu"
 /db_xref="taxon:4577"
 /clone="01S0518-03C1-H08"
 /clone_lib="UniformMu MutAIL Library"
 /note="Vector: TOPO-PCR4; DNA flanking Mu transposon insertions in Mu inactive lines were extracted from the UniformMu maize population by the thermo asymmetric interlaced PCR (TAIL) protocol using primers specific for the Mu terminal inverted repeat and a set of 16 arbitrary primers. Amplicons were size enriched using Sepharose 400 spin columns and cloned into the TOPO PCR4 vector."

ORIGIN
 Query Match 2.6%; Score 54.8; DB 29; Length 351;
 Best Local Similarity 58.6%; Pred. No. 0.00033;
 Matches 95; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 388 ATTGACGGTATTGAGTTATCAGGGGCGGATGTCACACTGTATGGCTCTGATGGGATG 447
 DB 182 ATTTGAACGTATTGAAGTTCTGCTGCGGCACTGCGGTTATGGCAACGGCGCGCG 123
 QY 448 GCGGGTGTGTGAATATCATTACCAAGAAATGCAGACAAATGGCTCTCTCCGTCAT 507
 DB 122 GCGGGGTGTGTTAAATCAATACCAAAAGGAGCGGAGTGGCACGGCTCTGGGAC 63
 QY 508 GCAGGGCTGAATCTGAGGAAGCAACAATGGGTAAACAGC 549
 DB 62 GCATATTTCAATCGCCAGAACATAAGAGGAAGGTGCCACC 21

RESULT 12
CB287407
LOCUS CB287407
DEFINITION CB287407.1 GI:28578361
ACCESSION CB287407
VERSION CB287407.1
KEYWORDS EST.
SOURCE Sus scrofa (pig)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

REFERENCE 1 (bases 1 to 598)
AUTHORS Dvorak, C.M.T., Hyland, K.A., Zhang, Y., Fahrenkrug, S.C. and Murtaugh, M.P.
TITLE Porcine jejunal Peyer's patch expressed sequences
JOURNAL Unpublished (2003)
COMMENT Contact: Murtaugh, M.P.

CB287407 598 bp mRNA linear EST 27-FEB-2003
 CMD59_C09_66 UMNMPM3 Sus scrofa cDNA clone PPSUBLB_44C09_5', mRNA sequence.

[illegible][illegible]

and submitted by the Animal Biotechnology Center (URL: <http://primer.ansci.umn.edu/software.html>), University of Minnesota. Trace files have been submitted to the NCBI trace archive (<http://www.ncbi.nlm.nih.gov/Traces/trace.cgi?>). Chromatograms were analyzed with Phred Version: 0.000925.c . Vector (pCMV_SPORT6, Invitrogen), Sus scrofa mitochondrion (GCMV_SPORT6|ref|NC_000845.1|, Escherichia coli K12|gi|5835862|ref|NC_000913.1|), Porcine reproductive and respiratory syndrome virus (NC_000913.1|gb|AF303357.1|AF303357), Altheromonas sp.16S rRNA (gi|11878202|gb|AF315322.1|AF315322), Sus scrofa 28S ribosomal gene (gi|14218471|emb|Y15322.1|Y15322.1), Sus scrofa 18S ribosomal RNA (gi|3388170|gb|AF080393.1|AF080393), and Sus scrofa 18S ribosomal RNA (gi|38855982|gb|AF102857.1|AF102857). Sequences were identified and masked by Crossmatch -minscore=20, -minmatch=12, -penalty=-8. SWAT alignment scores were set to match score=1, gap initial penalty=3 and gap extension penalty=-2. A continuous stretch of at least 50 bases of a PHRED quality 20 were required for submission.5, and 3' TERMINI were defined by a continuous stretch of 10 nucleotides with a PHRED quality not less than 20. A terminal stretch of at most 40 un-masked bases were automatically trimmed when flanking masked sequence. Sequences with an INTERNAL continuous stretch of at most 20 bases with PHRED quality less than 20 were automatically prepared for submission. HQUAL_START and HQUAL_STOP refer to the coordinates on CLIP_QUALITY and CLIP_VECTOR at the NCBI trace archive.

Plate: 44	row: C	column: 09
-----------	--------	------------

```

NCBI trace archive.
Plate: 44 row: C column: 09
Seq primer: SP6 primer
High quality sequence start: 66
High quality sequence stop: 277.
Location/Qualifiers
1..558
/organism="Sus scrofa"
/mol_type="mrna"
/db_xref="taxon:9823"
/clone="PPSUBLIB 44C09"
/tissue_type="Peyer's patch"
/clone_lib="UMNNP3"

```

```

1. 598
   /organism="Sus scrofa"
   /mol_type="rRNA"
   /db_xref="taxon:9823"
   /clone="PPSUBLIB_44C09"
   /tissue_type="Peyer's patch"
   /clone_lib="UMNNPM3"
   /ncntc="Organ: small intestine"

```


processed analyzed, groomed and submitted by the Animal Biotechnology Center (URL: <http://primer.ansci.umn.edu/software.html>), University of Minnesota. Trace files have been submitted to the NCBI trace archive (<http://www.ncbi.nlm.nih.gov/traces/trace.cgi?>). Chromatograms were analyzed with Phred version: 0.000925.c . Vector (PCW 5835862, Invitrogen) , Sus scrofa mitochondrion (gi|5835862|ref|NC_000845.1|), Escherichia coli k12 (gi|5835862|ref|NC_000913.1|), Porcine reproductive and respiratory syndrome virus (gi|11878202|gb|AF303357.1|AF303357), Alteromonas sp.16S rRNA gene (gi|4218471|emb|J15322.1|PX15322), Sus scrofa 28S ribosomal RNA (gi|3388170|gb|AF080393.1|AF080393), and Sus scrofa 18S ribosomal RNA (gi|3985982|gb|AF102857.1|AF102857). Sequences were identified and masked by Crossmatch -minscore=20, -minmatch=12, -penalty=-8. SWAT alignment scores were set to match score=1, gap initial penalty=-3 and gap extension penalty=-2. A continuous stretch of at least 50 bases of a PHRED quality 20 were required for submission. 5' and 3' TERMINI were defined by a continuous stretch of 10 nucleotides with a PHRED quality not less than 20. A terminal stretch of at most 40 un-masked bases were automatically trimmed when flanking masked sequence. Sequences with an INTERNAL_STOP stretch of at most 20 bases with PHRED quality less than 20 were automatically prepared for submission. HQUAL_START and HQUAL_STOP refer to the coordinates on CLIP_QUALITY and CLIP_VECTOR at the NCBI trace archive.

NCBI trace archive.
Plate: 29 row: G column: 05

Plate: 29 Row: G Column: 03
Seq primer: SP6 primer
High quality sequence start: 58
High quality sequence stop: 245.

FEATURES

```
1.609
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/clone="PPSUBLIB_29G05"
/tissue_type="Peyer's patch"
/clone_lib="MIMM003"
```

/clone.lib="CMNNP.M3"
 /note="Organ: jejunum; Vector: pCMVSPORT6; Site 1: NotI; Site 2: SalI; Jejunal Peyer's patches were isolated from healthy, 4-6 week old cross-bred pigs. RNA was extracted either immediately or after 3 hours stimulation in an Ussing chamber with one of the following treatments: 1) no treatment, 2) Salmonella choleraesuis vaccine strain SC-54, 3) phorbol myristate acetate, concanavalin A, and 8-bromo-cyclic AMP, and 4) lipopolysaccharide and cholera toxin. Each treatment was performed in the presence and absence of cycloheximide. Purified poly A + RNA from each of the treated tissues (2-4) was combined together, reverse transcribed, and cloned in to pCMVSPORT6 to make a library of approximately 530,000 recombinant clones with an average insert size of 1.0 kbp. Poly A+ RNA from freshly isolated and non-treated tissue from an Ussing chamber (treatment 1) was cloned in the same manner to produce an unstimulated cDNA library of approximately 900,000 clones with an average insert size of 1.5 kbp. Equal portions of the two libraries were pooled and then subtracted with porcine ST fibroblast RNA to create a subtracted porcine Peyer's patch library of approximately 6,000 clones with an average insert size of 1.0 kbp."

ORIGIN

ch	2.5%;	Score	51.6;	DB	14;	Length	609;
	li Similarity	70.4%;	Pred. No.	0.0039;			
	69; Conservative	0;	Mismatches	29;	Indels	0;	Gaps
175	CCCTCTGGCCGCCATTCAGCGTATTAGGTTATCAGGGGGCCGATGCCACACTGTATGG	434					
177	CCCTATTGCGGTTCTCCAGCGTGTGAATATATCCGTGGGGCCGCTCCGCTGTTATGG	236					
135	CTCTGATGCGATGGCGCGGTGTGGTGAATATCATTTACCA	472					
137	TTCGGATGCAATAGSCGGGGTGTGAATATCATCACGA	274					

RESULT 15	
CE286521	
LOCUS	
DEFINITION	
ACCESSION	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
COMMENT	

CB286521 618 bp mRNA linear EST 27-FEB-2003
 CMD39_B07_60 UMNPM3 Sus scrofa cDNA clone PFSUBLIB_24B07 5', mRNA
 sequence.
 CB286521
 CB286521 GI:28577475
 EST.
 Sus scrofa (pig)
 Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 1 (bases 1 to 618)
 Dvorak, C.M.T., Hyland, K.A., Zhang, Y., Fahrenkrug, S.C. and
 Murtaugh, M.P.
 porcine jejunal Peyer's patch expressed sequences
 Unpublished (2003)
 Contact: Murtaugh, M.P.
 Department of Veterinary Pathobiology
 University of Minnesota
 1971 Commonwealth Ave., St. Paul, MN 55108, USA
 Tel: 612-625-6735
 Fax: 612-625-5203

Email: murta001@umn.edu
 The PHRED quality scores of this sequence are supplied below: 39 39 39
 39 25 22 22 22 31 46 51 51 56 56 56 56 56 56 56 56 56 56
 56 56 51 51 51 51 51 56 56 56 51 51 35 34 32 32 34 34 35
 35 40 51 51 51 51 51 56 56 35 35 35 29 29 29 29 31 46
 35 35 38 35 35 40 45 43 43 40 40 40 37 40 45 45 40 40 40
 46 43 43 45 43 35 35 39 36 46 56 45 43 43 43 45 51 51
 56 35 35 35 35 35 42 46 46 56 56 51 51 43 43 43 41 41 56
 56 56 56 56 56 56 56 56 56 56 56 56 56 56 56 56 56 56
 56 56 56 42 41 41 45 45 45 51 51 43 40 40 38 38 38 56
 56 56 43 46 46 43 43 56 42 46 46 51 56 51 56 45 45 51 51
 56 51 45 45 45 45 45 48 46 43 46 56 56 56 56 56 56 51 51
 51 51 46 46 42 41 41 46 46 51 51 56 56 51 45 45 43 43 43
 36 36 36 40 40 41 51 45 45 45 45 45 45 45 45 45 45 45
 51 43 43 43 43 56 56 56 56 46 43 43 43 40 41 43 45 45 45
 45 51 56 56 56 56 56 51 45 45 45 43 43 43 39 36 32 36
 40 43 43 43 56 56 56 51 51 45 45 45 45 45 45 51 51 46 46
 56 43 56 56 56 56 56 43 43 43 43 43 43 43 45 43 40 56 43
 43 35 37 40 40 40 45 56 51 39 38 38 38 35 38 35 38 45
 45 45 51 51 51 51 43 46 45 35 35 35 35 35 45 45 51 56
 56 37 40 40 40 51 44 40 37 40 37 40 40 40 40 56 45 45
 45 45 51 45 45 40 37 37 37 37 40 40 37 35 39 40 40
 45 51 51 51 40 37 37 35 35 35 35 39 44 42 46 37 40
 40 37 40 35 40 37 40 40 45 56 56 42 43 42 37 37 40 40
 40 40 40 40 44 48 48 48 32 29 40 40 32 34 36 36 37
 36 44 32 32 29 25 15 20 19 20 22 29 29 20 19 27 32 29
 25 34 40 32 29 29 29 29 29 29 29 26 33 35 37 37 33
 29 29 29 29 40 29 34 48 34 25 22 17 20 25 24 20 22 22
 28 27 29 32 32 40 32 29 22 25 32 36 32 28 31 27
 Sequences were generated on ABI 3100 and 3700 capillary sequencers.
 Raw sequence data was batch processed analyzed, groomed and
 submitted by the Animal Biotechnology Center (University of
<http://primer.ansci.umn.edu/software1.html>), University of
 Minnesota. Trace files have been submitted to the NCBI trace
<http://www.ncbi.nlm.nih.gov/Traces/trace.cgi?>

archive (<http://www.ncbi.nlm.nih.gov/traces/trace.cgi>). Chromatograms were analyzed with Phred Version: 0.000925.c. Vector (pCMV SP8) and Invitrogen). Sus scrofa mitochondrion ref NC_000913.1, Escherichia coli K12 (g|5835862|g|5835862|ref NC_000845.1), Porcine reproductive and respiratory syndrome virus (g|11284847|emb|Y15322.1|PRY15322), Sus scrofa 28S ribosomal RNA (g|348180|gb|AF060393.1|AF060393), and Sus scrofa 18S ribosomal RNA (g|3885982|gb|AF102857.1|AF102857). Sequences were identified and masked by Crossmatch -m3score=20, -mismatch=12, -penalty=-8. SWAT alignment scores were set to match score=1, gap initial penalty=3 and gap extension penalty=2. A continuous stretch of at least 50 bases of a PHRED quality 20 were required for

submission.5' and 3' TERMINI were defined by a continuous stretch of 10 nucleotides with a PHRED quality not less than 20. A terminal stretch of at most 40 un-masked bases were automatically trimmed when flanking masked sequences. Sequences with an INTERNAL continuous stretch of at most 20 bases with PHRED quality less than 20 were automatically prepared for submission. HIQUAL START and HIQUAL STOP refer to the coordinates on CLIP_QUALITY and CLIP_VECTOR at the NCBI trace archive.

Plate: 24 row: B column: 07
Seq primer: SP6 primer
High quality sequence start: 57
High quality sequence stop: 265.
Location/Qualifiers
1. .618

FEATURES

source

/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/clone="PPSUBLIB_24B07"
/tissue_type="Peyer's patch"
/clone_lib="UMNMPM3"
/note="Organ: small intestine, jejunum; Vector: pCMVSPORT6; Site 1: NotI; Site 2: SalI; Jejunal Peyer's patches were isolated from healthy, 4-6 week old cross-bred pigs. RNA was extracted either immediately or after 3 hours stimulation in an Ussing chamber with one of the following treatments: 1) no treatment, 2) Salmonella choleraesuis vaccine strain SC-54, 3) phorbol myristate acetate, concanavalin A, and 8-bromo-cyclic AMP, and 4) lipopolysaccharide and cholera toxin. Each treatment was performed in the presence and absence of cycloheximide. Purified poly A + RNA from each of the treated tissues (2-4) was combined together, reverse transcribed, and cloned in to pCMVSPORT6 to make a library of approximately 530,000 recombinant clones with an average insert size of 1.0 kbp. Poly A+ RNA from freshly isolated and non-treated tissue from an Ussing chamber (treatment 1) was cloned in the same manner to produce an unstimulated cDNA library of approximately 900,000 clones with an average insert size of 1.5 kbp. Equal portions of the two libraries were pooled and then subtracted with porcine ST fibroblast RNA to create a subtracted porcine Peyer's patch library of approximately 6,000 clones with an average insert size of 1.0 kbp."

ORIGIN

Query Match 2.5%; Score 51.6; DB 14; Length 618;
Best Local Similarity 70.4%; Pred. No. 0.0039;
Matches 69; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
QY 375 CCTCTGGCGCCATTGACGTTATTCAGTTATCAGGGCGGCGATGCCACACTGTATGG 434
Db 177 CCTATTGGCTGTGCCAGCGTTGATATATCCGTGGCGCGCCTCCGCTGTTATGG 236
QY 435 CTCTGATGCGATGGCGGTGTGGTGAATATCATATTACCA 472
Db 237 TTCCGATGCAATAGCGGGGTGGTGAATATCATCACGA 274

Search completed: October 12, 2004, 13:40:42
Job time : 5412 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 12, 2004, 13:40:48 ; Search time 999 Seconds

(without alignments)
10611.075 Million cell updates/sec

Title: US-10-625-972-4

Perfect score: 2091

Sequence: 1 atgcgaataaccactctggc.....cgctgaactatcagttctga 2091

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3340653 seqs, 2534783454 residues

Total number of hits satisfying chosen parameters: 6681306

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*

2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*

3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*

4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*

5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*

6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*

7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*

8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*

9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*

10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*

11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*

12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*

13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*

14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*

15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*

16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*

17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*

18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

19: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	2089.4	99.9	87563	15	US-10-114-170-57
C 2	2079.8	99.5	76804	15	US-10-085-959-109
C 3	125.6	6.0	2046	15	US-10-238-075-1076
C 4	125.6	6.0	12943	15	US-10-238-075-1049
C 5	64.2	3.1	758	11	US-09-764-875-112
C 6	64.2	3.1	822	9	US-09-764-868-94
C 7	57.6	2.8	686	9	US-09-764-868-510
C 8	57.6	2.8	686	11	US-09-764-875-424
C 9	51.8	2.5	296	11	US-09-864-408A-851
C 10	51.4	2.5	2731748	17	US-10-238-075-526
C 11	48.6	2.3	1362	15	US-10-238-075-507
C 12	48.6	2.3	12264	15	US-10-085-959-11
C 13	48.6	2.3	23654	15	US-10-085-959-11
C 14	47	2.2	2760	13	US-10-194-163-462

15	43	2.1	2061	13	US-10-282-122A-12338	Sequence 12338, A
16	42	2.0	2182	13	US-10-194-163-916	Sequence 916, App
17	39.6	1.9	4334	13	US-10-194-163-619	Sequence 619, App
18	39.4	1.9	990	13	US-10-282-122A-12119	Sequence 12119, A
19	39.2	1.9	699	15	US-10-184-644-138	Sequence 138, App
20	39.2	1.9	699	15	US-10-184-634-138	Sequence 389, App
21	38.8	1.9	2127	16	US-10-264-237-389	Sequence 126, App
22	38.8	1.9	4758	15	US-10-175-523-126	Sequence 12731, A
23	38.4	1.8	2052	13	US-10-282-122A-12731	Sequence 12007, A
24	38	1.8	576	15	US-10-029-386-12007	Sequence 24900, A
25	38	1.8	674	15	US-10-029-386-24900	Sequence 20342, A
26	37.6	1.8	683	13	US-10-142-426-394	Sequence 394, App
27	37.6	1.8	1184	15	US-10-123-155-394	Sequence 394, App
28	37.6	1.8	1184	15	US-10-146-731-394	Sequence 394, App
29	37.6	1.8	1184	15	US-10-140-472-394	Sequence 394, App
30	37.6	1.8	1184	15	US-10-141-761-394	Sequence 394, App
31	37.6	1.8	1184	15	US-10-142-885-394	Sequence 394, App
32	37.6	1.8	1184	15	US-10-158-790-394	Sequence 394, App
33	37.6	1.8	1184	16	US-10-137-871-394	Sequence 394, App
34	37.6	1.8	1184	16	US-10-140-923-394	Sequence 394, App
35	37.6	1.8	1184	16	US-10-141-759-394	Sequence 394, App
36	37.6	1.8	1184	16	US-10-141-759-394	Sequence 394, App
37	37.6	1.8	1184	16	US-10-141-759-394	Sequence 394, App
38	37.6	1.8	1184	16	US-10-140-805-394	Sequence 394, App
39	37.6	1.8	1184	16	US-10-140-864-394	Sequence 394, App
40	37.2	1.8	2139	15	US-10-238-075-262	Sequence 22619, A
41	37.2	1.8	2364	13	US-10-282-122A-22619	Sequence 361, App
42	37.2	1.8	2890	9	US-09-881-752A-361	Sequence 152, App
43	37.2	1.8	7306	15	US-10-085-959-152	Sequence 261, App
44	37.2	1.8	7315	15	US-10-238-075-261	Sequence 5399, App
45	36.8	1.8	300	11	US-09-864-408A-5399	

ALIGNMENTS

RESULT 1

US-10-114-170-57/c
; Sequence 57, Application US/10/114/170
; Publication NO. US20030023075A1
; GENERAL INFORMATION:

APPLICANT: Blattner, Frederick R.

Perna, Nicole T.

Plunkett, Guy

Welch, Rod

TITLE OF INVENTION: No. US20030023075A1el Sequences of E. coli O157

NUMBER OF SEQUENCES: 265

CORRESPONDENCE ADDRESS:

ADDRESSEE: Quarles & Brady

STREET: 1 South Pinckney Street

CITY: Madison

STATE: WI

COUNTRY: US

ZIP: 53701-2113

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Word Perfect 8.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/114,170

FILING DATE: 01-Apr-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/453,702

FILING DATE: 03-DEC-1999

APPLICATION NUMBER: 60/110,955

FILING DATE: 04-DEC-1998

ATTORNEY/AGENT INFORMATION:

NAME: Seay, Nicholas J.

REGISTRATION NUMBER: 27386

REFERENCE/DOCKET NUMBER: 960296.95017

TELECOMMUNICATION INFORMATION:									
TELEPHONE: (608) 251-5000									
TELEFAX: (608) 251-9166									
INFORMATION FOR SEQ ID NO: 57:									
SEQUENCE CHARACTERISTICS:									
LENGTH: 87563									
TYPE: nucleic acid									
STRANDEDNESS: double									
TOPOLOGY: linear									
MOLECULE TYPE: DNA (genomic)									
SEQUENCE DESCRIPTION: SEQ ID NO: 57:									
US-10-114-170-57									
Query Match 99.9%; Score 2089.4; DB 15; Length 87563;									
Best Local Similarity 100.0%; Pred. No. 0;									
Matches 2090; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
QY	1	ATGCGAATAACCACTCTGGCTTCGGTAGTCAATCCCTGTCTCGGATTTTCAGCCAGCAGC	60						
DB	49991	ATGCGAATAACCACTCTGGCTTCGGTAGTCAATCCCTGTCTCGGATTTTCAGCCAGCAGC	49932						
QY	61	ATAGCTGCTGCAGAGATGTGATGTTCTCGGCATCCGGCTATGAGAAAAGCTGACT	120						
DB	49931	ATAGCTGCTGCAGAGATGTGATGTTCTCGGCATCCGGCTATGAGAAAAGCTGACT	49872						
QY	121	AACGCCAGCCGACAGTGTCTGTGATTAGCCAGGAGGAATTCAGTCCAGCCAGTACCAC	180						
DB	49871	AACGCCAGCCGACAGTGTCTGTGATTAGCCAGGAGGAATTCAGTCCAGCCAGTACCAC	49812						
QY	181	GATCTGGCGAGGCTCTGAGATCAGTAGAGGTGTGATGTTGAAAGTGTGACGGTAAA	240						
DB	49811	GATCTGGCGAGGCTCTGAGATCAGTAGAGGTGTGATGTTGAAAGTGTGACGGTAAA	49752						
QY	241	ACCGAGGCTCGAATCAGCATCCAGGAATCCAGCAATTCAGTCCAGCCAGTACCAC	300						
DB	49751	ACCGAGGCTCGAATCAGCATCCAGGAATTCAGCAATTCAGTCCAGCCAGTACCAC	49692						
QY	301	GATGTGTTCGTCAAGCGGAGAGAGTGAAGTGAATCCCAACGGTTCCTGCAATGAAT	360						
DB	49691	GATGTGTTCGTCAAGCGGAGAGAGTGAAGTGAATCCCAACGGTTCCTGCAATGAAT	49632						
QY	361	ACCGGTTTCATGCCCTCTGCGCCCATGAGCTGATGAGTTATCAGGGGCGCGATG	420						
DB	49631	ACCGGTTTCATGCCCTCTGCGCCCATGAGCTGATGAGTTATCAGGGGCGCGATG	49572						
QY	421	TCCACACTGTATGGCTCTGATCGATGGCGGTGTGGTGAATATCATTAACAGAAAGAT	480						
DB	49571	TCCACACTGTATGGCTCTGATCGATGGCGGTGTGGTGAATATCATTAACAGAAAGAT	49512						
QY	481	GCAGCAATGCTCTCTTCGTCAATGCAGGGCTGAATCTGCAGGAAGCAACAATGG	540						
DB	49511	GCAGCAATGCTCTCTTCGTCAATGCAGGGCTGAATCTGCAGGAAGCAACAATGG	49452						
QY	541	GCTAACAGCAGCCAGTTAAATTTCTGGAGCAGTGTCCCTTTGTGATGATTTCTGTGAGC	600						
DB	49451	GCTAACAGCAGCCAGTTAAATTTCTGGAGCAGTGTCCCTTTGTGATGATTTCTGTGAGC	49392						
QY	601	CTGCGAGTACCGGTAGACACAACAGCGTCAAGGTTTCATCGGTCAATCACTGAGCGAT	660						
DB	49391	CTGCGAGTACCGGTAGACACAACAGCGTCAAGGTTTCATCGGTCAATCACTGAGCGAT	49332						
QY	661	ACAGCAGGACCGCTATTCCTTATCCACGGAGTCCAGAGATTAATCTTGTGTCAGCT	720						
DB	49331	ACAGCAGGACCGCTATTCCTTATCCACGGAGTCCAGAGATTAATCTTGTGTCAGCT	49272						
QY	721	CTTGACTGAAAGCGCTCGAGCAGATGTCTGTGTTTGTATGATPACCCCGGAG	780						
DB	49271	CTTGACTGAAAGCGCTCGAGCAGATGTCTGTGTTTGTATGATPACCCCGGAG	49212						
QY	781	CGTTATGATAACCGGATGGGCAACTGGGGAGTCTGACGGGGGATATGACCGGACCTTG	840						
DB	49211	CGTTATGATAACCGGATGGGCAACTGGGGAGTCTGACGGGGGATATGACCGGACCTTG	49152						

789 TAAACGGGATGGGCAACTGGGAGTCTGACGGGGGGATGATGACGGGACCCCTGGCTATGA 848
 Db TGCCTGGACTATGCGGGAAATCTTAAACAAACCAACAGTAAAGAGAAACGCAATATC 848
 QY GCGAAACAAATTTTCACTGGCTATGATCATCTTTTCACTTCCGAAACATGGAAATCGTA 908
 Db ACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 902
 QY TCTGAACTGGAAACGAGACAGAAATAAAGTCTGAGCTTGTAGCGAGTGTACTGAAAGCG 968
 Db AATTGCTGTTTATCAGGAGAAAGTTATCTGAGAGTTAAATCAGGTAAAAAAGATAATA 962
 QY CGACAAATGGGGCTTCCCGCTGACGCGCGGAGCTTAAAGAAATCGAACTTATCTCTGAA 1028
 Db TAATCAATGGGATCTTAATACGAGTCAAGAAACCGGAAATTAACCAACAAATCATAGA 1022
 QY TTCAATTAAGTCTTACCCCTCTGGGAGATCTCATCTGTTTACGTTGGGGGCGAGTTTCA 1088
 Db TGCATAAGTACGCAATTTCTGCGGAAATGTACTGACCATCGAGGTCAATTTTCAGCA 1082
 QY GAGCTGCTCCATGAAGACGAGTGTCTTCCAGACAGGTTGAAA-----CTTTCGG 1142
 Db TGCAGAGCTCGGTGATGACTCAGCCACGGGTAAAGAAACGACAGAAACACAGTCTGTTT 1142
 QY GCAGAAAGCTGCTGCTGATTTTCTGAGGATGAGTGGCATCTCACGGATGCACTTGGCT 1202
 Db AATTAAACAGAAAGCTGTTTATAGAAATGATGATGATGATGATGATGATGATGATGATG 1202
 QY GACTCGGGGAGCGCTATGAACATCATGACAAATCGGGGACACTTCACTGCTGGCTGTC 1262
 Db GACTGGAGGAGCTGCTCGATATCATGAATCTATGCGAGTTACTGGAATCCAGATT 1262
 QY ATATCTGCTGGGATGTCGAGATGCTGAGCTGAAAGCGGTGTCACCGGATA 1322
 Db GTAGCTGTTTATACCTGACCGATATCTCACACTCAAGGGGGGATCGCAAGCATT 1322
 QY TAAGGCAACCAAGATGGGCGAGTACATATAAGGAATAGTGGTGTCTCGGGCAGGAAA 1382
 Db TCGGCTCTTCAATCTGAGGAGTGGTCTGATTTGGAACACTGACGAGGCTGCTG 1382
 QY AACAAATCTACTTGTAAACCCGAGCTGAACCCGGAAGAGCGTCACTGATGAGGCTGG 1442
 Db CTCTATTATGATGGAACAGGAGCTGAAACCCGAGACCGAGTGAACCGAAGAGATCGG 1442
 QY GGTGATTAACGATTAACCCCGGCTGATGATGCAATGTCACAGGTTTATGACTGACTT 1502
 Db TATTATTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1502
 QY CTCACAAAGATGCTCTTATTCATATAATGATAA 1538
 Db TAAAAATAAGTTGACCACTTACGATATAGGTACAAA 1538

RESULT 4
 US-10-238-075-1049
 ; Sequence 1049, Application US/10238075
 ; Publication No. US20030148324A1
 ; GENERAL INFORMATION:
 ; APPLICANT: I.N.S.E.R.M.
 ; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolated
 ; TITLE OF INVENTION: E.coli, and biological uses of these polynucleotides and of their
 ; FILE REFERENCE: BLANDINE
 ; CURRENT APPLICATION NUMBER: US/10/238,075
 ; CURRENT FILING DATE: 2002-09-10
 ; PRIORITY APPLICATION NUMBER: 0003145
 ; PRIOR FILING DATE: 2000-03-10
 ; NUMBER OF SEQ ID NOS: 1576
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 1049
 ; LENGTH: 12943
 ; TYPE: DNA
 ; ORGANISM: Escherichia coli
 US-10-238-075-1049

Query Match 6.0%; Score 125.6; DB 15; Length 12943;
 Best Local Similarity 46.3%; Pred. No. 1.3e-30;
 Matches 684; Conservative 0; Mismatches 759; Indels 33; Gaps 7;
 QY 72 AGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 131
 Db 10799 AGAAGATACGCTTGTGTTACTGCTCCGGTTCCTCAGCAGCTCAGAAATGCCCCGC 10858
 QY 132 CAGTGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 191
 Db 10859 CAGTGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 10918
 QY 192 GCCTCTGAGATCAGTAGAGGGTGTGATGTTTGAAGTGTGATGATGATGATGATGATGATGAT 251
 Db 10919 TG-----CAGTAAAGATGTTGAAGGATTAAGTATCACTGTGGGATGAAAC 10969
 QY 252 GGAATACAGATCCGAGGAAATCCAGCCAGTTTACAGCTGATGATGATGATGATGATGATGATGAT 311
 Db 10970 GGATATCAGTATACGTTGTTAAGTGGGATTAACAGCTGATTTCTGCTGATGAGACGAG 11029
 QY 312 TCAGGGCGGAAGCAGTACGCTGACTCCCAACGGTTTCTGCCATGATGATGATGATGATGATGATGAT 371
 Db 11030 TCAGAGCG--GTGCGGAAATCCAGAACCGCAGCGCGGTTTGAAGCCGGAATTTAT 11086
 QY 372 GCGCCCTCTGCGCGGCAATGAGCTGATGAGGTTTATCAGGGGGCGGATGTCACACTGTA 431
 Db 11087 CCTCTCTGTGAAGCAATTTGAACGATTTGAAGTGTGATGATGATGATGATGATGATGATGAT 11146
 QY 432 TGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 491
 Db 11147 TGTGTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 11206
 QY 492 GCTCTCTTCCGTCAATGCA---GGGCTGAATCTGAGGAAGCAACAAATGGGTTAAACAG 548
 Db 11207 ATGGGATGCGTACTTGGACTTGGGGGATTAATCAGGAACATGGAATTTGGTAACTC 11266
 QY 549 CAGCCAGTTTAAATTTCTGAGCAGTGTCTCCCTTGTGATGATGATGATGATGATGATGATGATGAT 608
 Db 11267 AACCAAAATGACTTCTATCTGTCAGGCCCAATGATTAAGGATAAACTTGGTCTTCAGCT 11326
 QY 609 ACGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 668
 Db 11327 ATATGAGGAATGAATCTGCAAGG-----AGATAGTATCTCTCAGGGA-ACACCGGC 11380
 QY 669 CAGCGCTATTTCTTATCCCAAGGATCAGAAATATATACTTTGGTGGACGCTTGTGACTG 728
 Db 11381 ABAAGATAATAAGAAATATAACGCAACGCTCCAGTTTACTCCGACTGAAAGCCAGAAAT 11440
 QY 729 GAAGCGCTCGAGCAGGATGCTCTGTTTGTATGATGATGATGATGATGATGATGATGATGATGAT 788
 Db 11441 TGTTTTGAATATGGAATAAATAACAGGTGATGATGATGATGATGATGATGATGATGATGAT 11500
 QY 789 TAACCGGGATGGCAACTGGGAGTCTGACGGGGGATATGACCGGACCCCTGCGCTATGA 848
 Db 11501 TGCTTGACTATGCGGGGAAATCTTAAACAAACACAGTAAAGAGAAACGCAATATTC 11560
 QY 849 GCGAAACAAATTTTCACTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 908
 Db 11561 ACCTAGTCACTGGGTAGCAGCAT-----GGAATGCCAGGGCGAAATATCTGCATCTGA 11614
 QY 909 TCTGACTGGAACGACACAGAAATAAAGTCTGTGAGCTTGTACGAGTGTACTGAAAGCG 968
 Db 11615 AATTGCTGTTTATCAGAGAAATGATTAATCTGTGAGGTTAATCAGGTAAAGAAATGATAA 11674
 QY 969 CGACAAATGGGGCTTCCCGGTCAGCCGGGGAGCTTAAGGAATCGAACTTATCTCTGAA 1028
 Db 11675 TAATCATTGGATCTTAATTTACAGTCAAGAAACCGGAAATAACCAACATCATAGA 11734
 QY 1029 TTCAATTAAGTCTTACCCCTCTGAGGAATCTCATCTGTTTACGTTGCGGGGGGAGTTTCA 1088
 Db 11735 TGCAAAATGACGGCAATTTCTGCCGAAATATGATGATGATGATGATGATGATGATGATGATGAT 11794

QY 1089 GAGCTCGTCCATGAAAGCGAGTGTCTCTGCCAGCACAGGTGAAA-----CTTTCCG 1142
Db 11795 TGCAGAGCTCCGATGATCTAGCCACCGGTAAATAACACAGAGAAACACAGCTCTGTTTC 11854
QY 1143 GCAGAAAGCTGGTTCGGTATTTGCTGAGGATGAGTGGCATCTCACGGATGCACTTGCCT 1202
Db 11855 AATTAAACAGAAAGCTGTTTTATAGAAAATGAATATGACGCAACGGATCTCTCGCCT 11914
QY 1203 GACTCGGCGGACCGCTATACACATCATGACAAATTCGGGACACATTCAGTCCCGTGC 1262
Db 11915 GACTGAGGACCTGCTGCTGAAATCATGAATCTATGCGACGTACTGGAATCCAGATT 11974
QY 1263 ATATCTGCTGCTGGATGTGGCAGATGCTGAGCGTGAAGGCGGTGTGACACCGGATA 1322
Db 11975 GTACGCTGTTTATAACCTGACCGATAATCTCACACTCAAGGGGGATCGCAAGAGCAT 12034
QY 1323 TAAGGACCCAGAAATGGGCGAGTACATAAGGATTTAGTGTGTCTGGGCGAGGAAA 1382
Db 12035 TCGGGCTCTTCAATTCGTGAGTGTGAGTCTCTGATTTGGAACACTGACGCGGTTGTC 12094
QY 1383 AACAAATCTACTTGGTAACCCGACCTGAAGCGGAGAGAGCGTCACTTATGAGGCTGG 1442
Db 12095 CTCATTATGATGAAACAGGACCTGAAACCGGAGACCACTGTAAACGAGAGATCGG 12154
QY 1443 GGTGTATACGATAACCCCGGCTGTGAATGCCAATGTACAGGTTTTATGACTGACTT 1502
Db 12155 TATTATTATAGTAATGATGATGTTTTCGGCGAGCGAGCTGTTTAACTACTGATT 12214
QY 1503 CTCACAAAGATTGCTCTTATTCATTAATGATAA 1538
Db 12215 TAAAAATAAGTTGACCAAGTTACGATATAGGTACAAA 12250

RESULT 5

US-09-764-875-112
; Sequence 112, Application US/09764875
; Publication No. US2004001896A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJ202
; CURRENT APPLICATION NUMBER: US/09/764,875
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1249
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 112
; LENGTH: 758
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-875-112

Query Match 3.1%; Score 64.2; DB 11; Length 758;
Best Local Similarity 57.1%; Pred. No. 2.4e-10;
Matches 117; Conservative 0; Mismatches 88; Indels 0; Gaps 0;
QY 388 ATTGACGCTATTGAGTTATCATAGGGGCGGATGTCACACTGTATGGCTCTGATCGCATG 447
Db 427 ATTGACGCTATTGAGTTATCATAGGGGCGGATGTCACACTGTATGGCTCTGATCGCATG 486
QY 448 GCGGCTGTGGTGAATATCATTAACCAAGAAAGATGAGCAAAATGGCTCTCTTCGGTCAAT 507
Db 487 GCGGCTGTGGTGAATATCATTAACCAAGAAAGATGAGCAAAATGGCTCTCTTCGGTCAAT 546
QY 508 GCAGGCTGAATCTCGAGAAAGCAAAATGGGTTACAGCAGCCAGTTAAATTTCTGG 567
Db 547 GCATATTTCAATGCGCCAGAAACATAAGAGGAAGGTGCCCAACCGCACTAACTTTAGC 606
QY 568 AGCAGTGTGCTCCCTTGTGGATGATT 592
Db 607 CTGACCGTCCGCTGGCGACGAAT 631

RESULT 6

US-09-764-868-94/c
; Sequence 94, Application US/09764868
; Patent No. US2002016871A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 94
; LENGTH: 822
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (795)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (804)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-868-94

Query Match 3.1%; Score 64.2; DB 9; Length 822;
Best Local Similarity 57.1%; Pred. No. 2.6e-10;
Matches 117; Conservative 0; Mismatches 88; Indels 0; Gaps 0;
QY 388 ATTGACGCTATTGAGTTATCATAGGGGCGGATGTCACACTGTATGGCTCTGATCGCATG 447
Db 324 ATTGACGCTATTGAGTTATCATAGGGGCGGATGTCACACTGTATGGCTCTGATCGCATG 265
QY 448 GCGGCTGTGGTGAATATCATTAACCAAGAAAGATGAGCAAAATGGCTCTCTTCGGTCAAT 507
Db 264 GCGGCTGTGGTGAATATCATTAACCAAGAAAGATGAGCAAAATGGCTCTCTTCGGTCAAT 205
QY 508 GCAGGCTGAATCTCGAGAAAGCAAAATGGGTTACAGCAGCCAGTTAAATTTCTGG 567
Db 204 GCATATTTCAATGCGCCAGAAACATAAGAGGAAGGTGCCCAACCGCACTAACTTTAGC 145
QY 568 AGCAGTGTGCTCCCTTGTGGATGATT 592
Db 144 CTGACCGTCCGCTGGCGACGAAT 120

RESULT 7

US-09-764-868-510
; Sequence 510, Application US/09764868
; Patent No. US2002016871A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 510
; LENGTH: 686
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (19)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (28)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-868-510


```

Query Match      2.8%; Score 57.6; DB 9; Length 686;
Best Local Similarity 58.0%; Pred. No. 4.2e-08;
Matches 102; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 388 ATTGAGCGTATTGAGTTATCAGGGGCGGATGCCACACTGTATGGCTCTGTATGCGATG 447
DB 499 ATTGAAGTATTGAGTTCTCGTGGTCCGCGACTGCGGTTATGCAACGGCGGCG 558

QY 448 GCGGCTGTGGTATATCATTTACAGAAAGAAATGCAGAAATGGCTCTTCTTCGTCAT 507
DB 559 GCGGCGTGTGTTAATCATTTACCAAAAAGGCGAGTGCGACCGCTCTCTGGGAC 618

QY 508 GCAGGCTGTATCTGCAGGAAGCAACAATGGGTTACAGCAGCCAGTTTAATTT 563
DB 619 GCATATTTCAATGCGCCAGAAACATAAAGAGAGAGTGCACCAACGCATCACTT 674

RESULT 8
US-09-764-875-424
; Sequence 424, Application US/09764875
; Publication No. US20040018969A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P0202
; CURRENT APPLICATION NUMBER: US/09/764,875
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1249
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 424
; LENGTH: 686
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (19)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (28)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-875-424

Query Match      2.8%; Score 57.6; DB 11; Length 686;
Best Local Similarity 58.0%; Pred. No. 4.2e-08;
Matches 102; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 388 ATTGAGCGTATTGAGTTATCAGGGGCGGATGCCACACTGTATGGCTCTGTATGCGATG 447
DB 499 ATTGAAGTATTGAGTTCTCGTGGTCCGCGACTGCGGTTATGCAACGGCGGCG 558

QY 448 GCGGCTGTGGTCAATATCATTTACCAAGAAAGAAATGCAGAAATGGCTCTTCTTCGTCAT 507
DB 559 GCGGCGTGTGTTAATCATTTACCAAAAAGGCGAGTGCGACCGCTCTCTGGGAC 618

QY 508 GCAGGCTGTATCTGCAGGAAGCAACAATGGGTTACAGCAGCCAGTTTAATTT 563
DB 619 GCATATTTCAATGCGCCAGAAACATAAAGAGAGAGTGCACCAACGCATCACTT 674

RESULT 9
US-09-864-408A-851
; Sequence 851, Application US/09864408A
; Publication No. US2004000947A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Shinkets, Richard A.
; TITLE OF INVENTION: No. US2004000947A1 Human Polynucleotides and Polypeptides Encod
; FILE REFERENCE: 21402-012
; CURRENT APPLICATION NUMBER: US/09/864,408A
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/206,690
; PRIOR FILING DATE: 2000-05-24

```

```

; NUMBER OF SEQ ID NOS: 9068
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 851
; LENGTH: 296
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-864-408A-851

Query Match      2.5%; Score 51.8; DB 11; Length 296;
Best Local Similarity 60.1%; Pred. No. 2.4e-06;
Matches 86; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 1372 GGGCAGGGGAAAAAATAATCTACTTGGTAACCCCGACCTGGAAGCCGGAAGAGAGCTCAGT 1431
DB 7 GGTGGGCGACCTCAATCATTTATCGGCAACCGGACCTGCAGCCAGAGAGAGCTGAGC 66

QY 1432 TATGAGGCTGGGGTGTATTACGATAACCCCGCGGCTCTGAATGCCAATGTACAGGTTT 1491
DB 67 GAAGAGATCGGCTGTATGTGGGATAACCAACAGGCGCTGAACGCGGCTAACCTGTTC 126

QY 1492 ATGACTGACTTCTCCAAACAAGAT 1514
DB 127 AACACCGACTTCAAGACAAGAT 149

RESULT 10
US-10-297-465A-1
; Sequence 1, Application US/10297465A
; Publication No. US20040142413A1
; GENERAL INFORMATION:
; APPLICANT: Simpson, Andrew
; APPLICANT: Reinach, Fernando
; APPLICANT: Setubal, Joao
; APPLICANT: Medanis, Joao
; APPLICANT: Arruda, Paulo
; TITLE OF INVENTION: Isolated Genome of Xylella fastidiosa and Uses Thereof
; FILE REFERENCE: PAPER 202 US (10213376)
; CURRENT APPLICATION NUMBER: US/10/297,465A
; CURRENT FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: PCT/IB01/01618
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: 60/209,906
; PRIOR FILING DATE: 2001-06-17
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 1
; LENGTH: 2731748
; TYPE: DNA
; ORGANISM: Xylella fastidiosa
US-10-297-465A-1

Query Match      2.5%; Score 51.4; DB 17; Length 2731748;
Best Local Similarity 54.0%; Pred. No. 0.00096;
Matches 128; Conservative 0; Mismatches 106; Indels 3; Gaps 1;

QY 240 AACCGAGGCTCGAAATCAGCATCCGAGGAATCCAGCCAGTTACAGCTGATCTGAT 299
DB 347778 AATGGGTGGAGTGGTATCTCGTTGGTGGCTCCAGCTCCACTCTGCTGCTGAT 347837
QY 300 TGATGGTGTTCGTGAGGCGGGAAGCAGTGCAGTCTCCCAACGGTTTTTCTGCCATGAA 359
DB 347838 TAACGG--TCGTGCTATCGTCTCTTACGGTTGGCCGACGATGGTCAGAAAGTTGTCG 347894

QY 360 TACCGGGTTCATCCCGCTCTGCGCCCATTTAGCGGTTATGAGTTATCAGGGGCGCGAT 419
DB 347895 TGATTTGAGTAGCATCCCGATGGATGCGGTGAGCGTGTGAAGTTCTGAAGACCGGTGC 347954
QY 420 GTCCACACTGTATGGTCTGTGATGCGATGGCGGCTGTGGTGAATATCATTTACAGAAA 476
DB 347955 GTCTCTCAATTTAGGATCTTGACCGGATTCCTGGTGTAGTCAATATCATTTTGGCTAA 348011

RESULT 11

```

```

RESULT 13
US-10-085-959-11
; Sequence 11, Application US/10085959
; Publication No. US20030165870A1
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; APPLICANT: Welch, Rodney A.
; APPLICANT: Burland, Valerie D.
; TITLE OF INVENTION: No. US20030165870A1el Sequence of E. Coli CFT073
; FILE REFERENCE: 960296.97648
; CURRENT APPLICATION NUMBER: US/10/085,959
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: 60/242,412
; PRIOR FILING DATE: 2000-10-19
; NUMBER OF SEQ ID NOS: 255
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 23654
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (12189)..(12189)
; OTHER INFORMATION:
; NAME/KEY: misc feature
; LOCATION: (12189)..(12189)
; OTHER INFORMATION: Unsure
; US-10-085-959-11

Query Match      2.3%; Score 48.6; DB 15; Length 23654;
Best Local Similarity 54.9%; Pred. No. 0.00046;
Matches 96; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY      378 TCTGGCGGCATGACCGCTATTGAGGTTTACGGGGCCGATGTCCACACTGTATGGCTC 437
Db
DQ      2433 TCGGGCACTCGTAAACGATCGAAGTTGTCGGCGGCCAACTCCAGTCTGTACGGCAG 2492
QY      438 TGATGCGATGGCGGTGTGGTGAATATCATTTACCAAGAAAGATGCAGACAAATGGCTCTC 497
Db
DQ      2493 TGGCGGCTGGGAGGTGTAGTGGACTTCAGAACTGCCGATGCAGCAGATTTTCTCCCC 2552
QY      498 TTCCGTCATGACGGCTGAATCTCAGAAACCAACAATGGGTAAACGAGC 552
Db
DQ      2553 CGGAGAGACAAACGGTTTAAGTCTGTGGGGAATATCGCCAGTGGTGACACAGC 2607

RESULT 14
US-10-194-163-462/c
; Sequence 462, Application US/10194163
; Publication No. US20020172976A1
; GENERAL INFORMATION:
; APPLICANT: Ross, Bruce Carter
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS POLYNUCLEOTIDES
; AND USES THEREOF
; NUMBER OF SEQUENCES: 1120
; CORRESPONDENCE ADDRESS:
; ADDRESS: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: PALO ALTO
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA: US/10/194,163
; FILING DATE: 04-NOV-2002
; CLASSIFICATION: <Unknown>

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Basu, Shantanu
; REGISTRATION NUMBER: 43,318
; REFERENCE/DOCKET NUMBER: 529282000101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-813-5995
; TELEFAX: 650-494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 462
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2760 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEetical: NO
; ANTI-SENSE: UNKNOWN
; ORIGINAL SOURCE:
; ORGANISM: PORPHYROMONAS GINGIVALIS
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1..2760
; SEQUENCE DESCRIPTION: SEQ ID NO: 462
US-10-194-163-462

Query Match      2.2%; Score 47; DB 13; Length 2760;
Best Local Similarity 57.4%; Pred. No. 0.00043;
Matches 105; Conservative 0; Mismatches 75; Indels 3; Gaps 1;

QY 388 ATTGAGCGTATTGAGGTATCAGGGGGCGGATGCCACACTGTATGGCTCTGATGCGATG 447
Db 1444 ATCGAACGTGGAGGTAGTACGTGGTGAGGATCGGCTTGTACGGTCTCTTCGTATT 1385
QY 448 GCGGTGTGGTGAATATCATTTACCAGAAAGAAATGCAGACAAATGGCTCTCTTCGTCAT 507
Db 1384 CGCGAGTGGTGAATATCATTCACAGAACTTCTCACAATTTTTCACAT---TCAAT 1328
QY 508 CGAGGCTGAATCTGCAGAAAGCAACAAATGGGGTACAGCAGCCAGTTTAATTTCTGG 567
Db 1327 GAATCTCTGAGCTTTACCGGTTTCAGCAAGCTGGATAACACAGCAACTTCAATGCCTCC 1268
QY 568 AGC 570
Db 1267 ATC 1265

RESULT 15
US-10-282-122A-12338
; Sequence 12338, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282.122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26

```

```

; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 12338
; LENGTH: 2061
; TYPE: DNA
; ORGANISM: Bacteroides fragilis
US-10-282-122A-12338

Query Match      2.1%; Score 43; DB 13; Length 2061;
Best Local Similarity 60.9%; Pred. No. 0.0084;
Matches 70; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 388 ATTGAGCGTATTGAGGTATCAGGGGGCGGATGCCACACTGTATGGCTCTGATGCGATG 447
Db 406 ATCGACCATATCGAATATAGTCAAGAGGCTTCTTGGCCCTCTATGTTCTCAGCGATG 465
QY 448 GCGGTGTGGTGAATATCATTTACCAGAAAGAAATGCAGACAAATGGCTCTCTTCCG 502
Db 466 GCGGAGTTCATCAACATCATTCAGCGTAAAGCCAAAAGAAATTCGAGGCTTCCG 520

Search completed: October 12, 2004, 18:30:26
Job time : 1007 secs

```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 12, 2004, 11:56:38 ; Search time 156 Seconds

(without alignments)
7438.476 Million cell updates/sec

Title: US-10-625-972-4

Perfect score: 2091
Sequence: 1 atcggaataaccactctggc.....cgctgaactatcagttctga 2091

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2091	100.0	2091	1	US-08-765-081-4
2	2091	100.0	2091	3	US-09-098-082-4
3	2091	100.0	2091	5	PCT-US95-06994-4
4	2091	100.0	8041	1	US-08-765-081-1
5	2091	100.0	8041	3	US-09-098-082-1
6	2091	100.0	8041	5	PCT-US95-06994-1
7	2089.4	99.9	87563	4	US-09-453-702B-57
8	127	6.1	2055	4	US-09-543-681A-736
9	105.4	5.0	2100	4	US-09-489-039A-962
10	105	5.0	1535	1	US-08-480-510-1
11	105	5.0	1535	5	PCT-US94-01780-1
12	96.8	4.6	2034	4	US-09-543-681A-3764
13	68.8	3.3	2280	4	US-09-489-039A-703
14	68.6	3.3	2495	4	US-09-668-113A-1
15	63.4	3.0	2388	4	US-09-489-039A-6223
16	61.8	3.0	1326	4	US-09-252-991A-15544
17	61.8	3.0	1437	4	US-09-252-991A-15295
18	61.2	2.9	439	1	US-08-480-510-10
19	61.2	2.9	439	5	PCT-US94-01780-10
20	61.2	2.9	517	1	US-08-480-510-9
21	61.2	2.9	517	5	PCT-US94-01780-9
22	61.2	2.9	534	1	US-08-480-510-8
23	61.2	2.9	534	5	PCT-US94-01780-8
24	56.4	2.7	2790	4	US-09-328-352-465
25	55.6	2.7	1116	4	US-09-252-991A-13319
26	55.6	2.7	2235	4	US-09-252-991A-12994
27	55.6	2.7	3486	4	US-09-252-991A-12477

28	54.4	2.6	2337	4	US-09-489-039A-1284	Sequence 1284, Appl	
29	51.6	2.5	1845	4	US-09-307-973A-1	Sequence 1, Appl	
30	50.2	2.4	933	4	US-09-252-991A-14912	Sequence 14912, A	
C	31	50.2	2.4	939	4	US-09-252-991A-14458	Sequence 14458, A
32	48.4	2.3	1844	4	US-09-307-973A-2	Sequence 2, Appl	
C	33	47.4	2.3	633	4	US-09-252-991A-8508	Sequence 8508, Ap
C	34	47.4	2.3	2712	4	US-09-252-991A-8406	Sequence 8406, Ap
C	35	47.4	2.3	3528	4	US-09-252-991A-8473	Sequence 8473, Ap
C	36	47	2.2	2760	4	US-09-221-017B-462	Sequence 462, App
C	37	46.8	2.2	1854	4	US-09-252-991A-14350	Sequence 14350, A
C	38	46.8	2.2	2943	4	US-09-252-991A-14267	Sequence 14267, A
C	39	45	2.2	300	4	US-09-252-991A-15342	Sequence 15342, A
40	43.8	2.1	1497	4	US-09-489-039A-5016	Sequence 5016, Ap	
C	41	43.4	2.1	7218	1	US-08-232-463-14	Sequence 14, Appl
C	42	42.8	2.0	2286	4	US-09-328-352-1816	Sequence 1816, Ap
C	43	42	2.0	2182	4	US-09-221-017B-916	Sequence 916, App
C	44	41.8	2.0	405	4	US-09-252-991A-14718	Sequence 14718, A
C	45	41.8	2.0	1437	4	US-09-252-991A-15032	Sequence 15032, A

ALIGNMENTS

RESULT 1
US-08-765-081-4
; Sequence 4, Application US/08765081
; Patent No. 5798260
; GENERAL INFORMATION:
; APPLICANT: Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
; TITLE OF INVENTION: Escherichia Coli 0157:H7 Epithelial Adhesin
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
; STREET: 2800 Pacific First Center, 1420 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101-2347

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette-3.5 inch, 1.44Mb storage
COMPUTER: IBM PC compatible/Pentium
OPERATING SYSTEM: MS-Windows 3.1
SOFTWARE: Word for Windows-6.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/765,081
FILING DATE: March 26, 1997

PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/06994
FILING DATE: June 7, 1995

APPLICATION NUMBER: US 08/265,714
FILING DATE: June 24, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Shelton, Dennis K.

REGISTRATION NUMBER: 26,997
REFERENCE/DOCKET NUMBER: CHOR-1-10286
TELECOMMUNICATION INFORMATION:
TELEPHONE: 1-206-682-8100; 1-206-224-0718 (direct)

TELEFAX: 1-206-224-0779
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2091 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
DESCRIPTION: Corresponds to SEQ ID NO:1,
DESCRIPTION: nucleotides 3036-5126
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli 0157: H7
STRAIN: 86-24 NAIR
FEATURE:

; NAME/KEY: CDS									
; LOCATION: 1..2088									
US-08-765-081-4									
Query Match 100.0%; Score 2091; DB 1; Length 2091;									
Best Local Similarity 100.0%; Pred. No. 0;									
Matches 2091; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	ATCGGAATACCACTCTGGCTTCGGTAGTCATCCCTGTCTCGAATTTTCAGCCAGCAGC	60	Db	961	CTGAAGCGCGCAAAATGGGGCTTCGGGTGAGCGGGAGCTTAAAGGAATCGAACCTT	1020		
Db	1	ATCGGAATACCACTCTGGCTTCGGTAGTCATCCCTGTCTCGAATTTTCAGCCAGCAGC	60	QY	1021	ATCCCTGAATTCATTTACTGCTTACCCCTCTGGGAGAACTCTCATCTGGTTACGGTGGGGGG	1080		
Db	1	ATCGGAATACCACTCTGGCTTCGGTAGTCATCCCTGTCTCGAATTTTCAGCCAGCAGC	60	Db	1021	ATCCCTGAATTCATTTACTGCTTACCCCTCTGGGAGAACTCTCATCTGGTTACGGTGGGGGG	1080		
QY	61	ATAGCTGCTCGAGAGATGTGATGTTGCTCGGCATCGGCTATGAGAAAGAGCTGACT	120	QY	1081	GAGTTTCAGAGCTCGTCATGAAAGACGGAGTTGTCTCCCTCCAGCACAGGTGAAACTTTTC	1140		
Db	61	ATAGCTGCTCGAGAGATGTGATGTTGCTCGGCATCGGCTATGAGAAAGAGCTGACT	120	Db	1081	GAGTTTCAGAGCTCGTCATGAAAGACGGAGTTGTCTCCCTCCAGCACAGGTGAAACTTTTC	1140		
QY	121	AACGACGCCAGCTGATGTTCTCTGATTAAGCCAGGAGAAATTCAGCTCCAGCCAGTACCAC	180	QY	1141	CGGCAAGAAAGCTGGTGGTATTTGCTGAGGATGAGTGGCATCTCACGGATGACCTTTGG	1200		
Db	121	AACGACGCCAGCTGATGTTCTCTGATTAAGCCAGGAGAAATTCAGCTCCAGCCAGTACCAC	180	Db	1141	CGGCAAGAAAGCTGGTGGTATTTGCTGAGGATGAGTGGCATCTCACGGATGACCTTTGG	1200		
QY	181	GATCTCGCGAGGCTCTGAGATCACTAGAGGGTGTGGATGTTGAAAGTGTACGGGTAAA	240	QY	1201	CTGACTCGGGCAGCCGCTATGAAACATCATGAGCAATTCGGGGGACACTTTCAGTCCGCGT	1260		
Db	181	GATCTCGCGAGGCTCTGAGATCACTAGAGGGTGTGGATGTTGAAAGTGTACGGGTAAA	240	Db	1201	CTGACTCGGGCAGCCGCTATGAAACATCATGAGCAATTCGGGGGACACTTTCAGTCCGCGT	1260		
QY	241	ACCGAGGGCTCGAAATCAGCATCCGAGGAATGCCAGCAGTTACAGCGTATACTGATT	300	QY	1261	GCATATCTGGTCTGGGATGTGGAGATGCTGAGCGCTGAAAGCGGTGTGACACCGGA	1320		
Db	241	ACCGAGGGCTCGAAATCAGCATCCGAGGAATGCCAGCAGTTACAGCGTATACTGATT	300	Db	1261	GCATATCTGGTCTGGGATGTGGAGATGCTGAGCGCTGAAAGCGGTGTGACACCGGA	1320		
QY	301	GATGGTTCGTCAGGGCGGAAGCAGTGCAGTGCATCCCAACGGTTTTTCTGCCATGAAT	360	QY	1321	TATAAGGACCCAGAAATGGGGCAGCTACATAAAAGGGATTAAGTGGTGTCTCCGGCAGGGA	1380		
Db	301	GATGGTTCGTCAGGGCGGAAGCAGTGCAGTGCATCCCAACGGTTTTTCTGCCATGAAT	360	Db	1321	TATAAGGACCCAGAAATGGGGCAGCTACATAAAAGGGATTAAGTGGTGTCTCCGGCAGGGA	1380		
QY	361	ACCGGGTTCAATCCCTCTGCGCCCAATCAGAGCTATGAGGTTATCAGGGGCGCGATG	420	QY	1381	AAAAAAATCTACTTGTGTAAACCCGACCTGAAAGCCGGAAGAGCGTCAAGTTATGAGGCT	1440		
Db	361	ACCGGGTTCAATCCCTCTGCGCCCAATCAGAGCTATGAGGTTATCAGGGGCGCGATG	420	Db	1381	AAAAAAATCTACTTGTGTGTAAACCCGACCTGAAAGCCGGAAGAGCGTCAAGTTATGAGGCT	1440		
QY	421	TCACACTGTATGGCTCTGATGCGATGGCGGCTGAGTGAATATCATTAACAGAAAGAT	480	QY	1441	GGGGTGTATTACGATAACCCCGCGGTCTGAAATGCCAAATGTCTACAGGTTTTTATGACTGAC	1500		
Db	421	TCACACTGTATGGCTCTGATGCGATGGCGGCTGAGTGAATATCATTAACAGAAAGAT	480	Db	1441	GGGGTGTATTACGATAACCCCGCGGTCTGAAATGCCAAATGTCTACAGGTTTTTATGACTGAC	1500		
QY	481	CGACAGAAATGGCTCTCTCCGTCAATCGAGGGCTGAATCTGCAGGAAGACCAAAATGG	540	QY	1501	TTCTCCAAACAGATTTGCTCTTATTCATAAATGATAACCAATAGCTATGTAAACACAGC	1560		
Db	481	CGACAGAAATGGCTCTCTCCGTCAATCGAGGGCTGAATCTGCAGGAAGACCAAAATGG	540	Db	1501	TTCTCCAAACAGATTTGCTCTTATTCATAAATGATAACCAATAGCTATGTAAACACAGC	1560		
QY	541	GGTAACAGAGCCAGTTTAAATTTCTGGAGCAGTGGTCCCTGTGGATGATTTCTGCAGC	600	QY	1561	GGAAAGCCCGGTTGCAAGGTTGGAATTTGCCGACCAATTTGCCCTGTGGTCAAGGAT	1620		
Db	541	GGTAACAGAGCCAGTTTAAATTTCTGGAGCAGTGGTCCCTGTGGATGATTTCTGCAGC	600	Db	1561	GGAAAGCCCGGTTGCAAGGTTGGAATTTGCCGACCAATTTGCCCTGTGGTCAAGGAT	1620		
QY	601	CTGCAAGTACGGGTAGCACAAACCGGTCAGGGTTTCATCGGTCACATCACTGAGCGAT	660	QY	1621	GTCACGCTGTCACTGAAATTCACCTGGACCCGAAAGTGAACAAAGTGGTGAATACAAA	1680		
Db	601	CTGCAAGTACGGGTAGCACAAACCGGTCAGGGTTTCATCGGTCACATCACTGAGCGAT	660	Db	1621	GTCACGCTGTCACTGAAATTCACCTGGACCCGAAAGTGAACAAAGTGGTGAATACAAA	1680		
QY	661	ACAGCAGGACCGGTATTTCTTATCCACGAGGTCAAGAAATTAATCTTGGTGCACGT	720	QY	1681	GGTGGCGCGTGAATTAACCCCTGAAACATGGTGAATGCGGAAATGAACTGAGTGGCAGATC	1740		
Db	661	ACAGCAGGACCGGTATTTCTTATCCACGAGGTCAAGAAATTAATCTTGGTGCACGT	720	Db	1681	GGTGGCGCGTGAATTAACCCCTGAAACATGGTGAATGCGGAAATGAACTGAGTGGCAGATC	1740		
QY	721	CTTGACTGGAAGGGTTCGAGCAGGATGTGCTCTGGTTTGAATGAGTATGAGTATGAGT	780	QY	1741	ACCGAAGAGTGCATCATGGCTGGGTGGCGGTATTCGGGGGAAACACCAAGTTTCACC	1800		
Db	721	CTTGACTGGAAGGGTTCGAGCAGGATGTGCTCTGGTTTGAATGAGTATGAGTATGAGT	780	Db	1741	ACCGAAGAGTGCATCATGGCTGGGTGGCGGTATTCGGGGGAAACACCAAGTTTCACC	1800		
QY	781	CGTTATGATAACCGGGATGGCAACTGGGAGTCTGACGGGGGATATGACCGGACCTTG	840	QY	1801	CAGAAATTTTCGTCACTGAGCGCTGPAACAGAAAGTGTATGATGAGAAAGGAGAAATAC	1860		
Db	781	CGTTATGATAACCGGGATGGCAACTGGGAGTCTGACGGGGGATATGACCGGACCTTG	840	Db	1801	CAGAAATTTTCGTCACTGAGCGCTGPAACAGAAAGTGTATGATGAGAAAGGAGAAATAC	1860		
QY	841	CGCTATGAGCGAAACAAATTTTACGTGGCTATGATCATATCTTACCTTCACCTTCGAA	900	QY	1861	CTGAAAGCCCTGGACCGTGGTGGATGCAAGTCTGTGTGGAAAGTGTATGATGAGAAAGGAGAAATAC	1920		
Db	841	CGCTATGAGCGAAACAAATTTTACGTGGCTATGATCATATCTTACCTTCACCTTCGAA	900	Db	1861	CTGAAAGCCCTGGACCGTGGTGGATGCAAGTCTGTGTGGAAAGTGTATGATGAGAAAGGAGAAATAC	1920		
QY	901	AAATCGTATCTGAATGGAACGACAGAAATAAAGGTCGTGAGCTGTGACGCTGTA	960	QY	1921	CT			

Dd	2041	TATGTGATACCTAGCGGAATAATTACTGGATTCGCTGAACATCACTTCTGA	2091
RESULT 2			
US-09-098-082-4			
; Sequence 4, Application US/09098082			
; Patent No. 6040421			
; GENERAL INFORMATION:			
; APPLICANT: Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.			
; TITLE OF INVENTION: Escherichia Coli O157:H7 Epithelial Adhesin			
; NUMBER OF SEQUENCES: 12			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: Christensen, O'Connor, Johnson and Kindness			
; STREET: 2800 Pacific First Center, 1420 Fifth Avenue			
; CITY: Seattle			
; STATE: Washington			
; COUNTRY: USA			
; ZIP: 98101-2347			
; COMPUTER READABLE FORM:			
; MEDIUM TYPE: Diskette-3.5 inch, 1.44Mb storage			
; COMPUTER: IBM PC compatible/Pentium II			
; OPERATING SYSTEM: MS-Windows 95			
; SOFTWARE: Word for Windows-6.0			
; CURRENT APPLICATION DATA:			
; APPLICATION NUMBER: US/09/098,082			
; FILING DATE:			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: US 08/765,081			
; FILING DATE: March 26, 1997			
; APPLICATION NUMBER: PCT/US95/06994			
; FILING DATE: June 7, 1995			
; APPLICATION NUMBER: US 08/265,714			
; FILING DATE: June 24, 1994			
; ATTORNEY/AGENT INFORMATION:			
; NAME: Sheiness, Diana K.			
; REGISTRATION NUMBER: 35,356			
; REFERENCE/DOCKET NUMBER: CHOR-1-12402			
; TELECOMMUNICATION INFORMATION:			
; TELEPHONE: 1-206-682-8100; 1-206-224-0735 (direct)			
; TELEFAX: 1-206-224-0779			
; INFORMATION FOR SEQ ID NO: 4:			
; SEQUENCE CHARACTERISTICS:			
; LENGTH: 2091 base pairs			
; TYPE: nucleic acid			
; STRANDEDNESS: single			
; TOPOLOGY: linear			
; MOLECULE TYPE: DNA (genomic)			
; DESCRIPTION: Corresponds to SEQ ID NO:1,			
; DESCRIPTION: nucleotides 3036-5126			
; HYPOTHETICAL: NO			
; ANTI-SENSE: NO			
; ORIGINAL SOURCE:			
; ORGANISM: Escherichia coli O157: H7			
; STRAIN: 86-24 NALR			
; FEATURE:			
; NAME/KEY: CDS			
; LOCATION: 1..2088			
US-09-098-082-4			
Query Match 100.0%; Score 2091; DB 3; Length 2091;			
Best Local Similarity 100.0%; Pred. No. 0;			
Matches 2091; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
OY	1	ATGCGAATAAACCATCTGGCTTCGGTAGTCATCCCTGTCTCGGATTTTCAGCCAGCAGC	60
Dd	1	ATGCGAATAAACCATCTGGCTTCGGTAGTCATCCCTGTCTCGGATTTTCAGCCAGCAGC	60
OY	61	ATAGCTGTCGAGAGGATGCATGATGTCCTCGGCATCCGGCTATGAGAAAAGCTGACT	120
Dd	61	ATAGCTGTCGAGAGGATGCATGATGTCCTCGGCATCCGGCTATGAGAAAAGCTGACT	120
OY	121	AACGACGCCGACAGTGTTCTGTGATTAGCCAGGAGGAATTGCAGTCCAGCCAGTACCAC	180


```

QY 1261 GCATATCTGGTCTGGGATGTGGCAGATGCTGACCGTGAAGCGCGGTGTGACACGGGA 1320
Db 1261 GCATATCTGGTCTGGGATGTGGCAGATGCTGACCGTGAAGCGCGGTGTGACACGGGA 1320
QY 1321 TATAAGCACCAGATGCGGCGAGCTACATAAAGGATTTAGTGGTGTCTCGGCGCAGGA 1380
Db 1321 TATAAGCACCAGATGCGGCGAGCTACATAAAGGATTTAGTGGTGTCTCGGCGCAGGA 1380
QY 1381 AAAACAAATCTACTTGTGTAACCCGACCTCAAGCCGAGAGAGCTCAGTTTATGAGGT 1440
Db 1381 AAAACAAATCTACTTGTGTAACCCGACCTCAAGCCGAGAGAGCTCAGTTTATGAGGT 1440
QY 1441 GGGGTGTTACGATTAACCCCGCGGTCTGAATGCCAATGTACAGTTTATGAGGTAC 1500
Db 1441 GGGGTGTTACGATTAACCCCGCGGTCTGAATGCCAATGTACAGTTTATGAGGTAC 1500
QY 1501 TTCTCCAAACAGATTTGTCTTATTCATAAATGATAACCAATAGCTATGTAAACAGC 1560
Db 1501 TTCTCCAAACAGATTTGTCTTATTCATAAATGATAACCAATAGCTATGTAAACAGC 1560
QY 1561 GGAAGGCCCGGTGCAAGCGTGGAAATTTGCCGACATTTGCCGCTGGTGGTCAAGAGT 1620
Db 1561 GGAAGGCCCGGTGCAAGCGTGGAAATTTGCCGACATTTGCCGCTGGTGGTCAAGAGT 1620
QY 1621 GTCAAGCTGTCACTGAATTAACCTTGACCTGGACCCGAGTGAACAGCTGATGGTATACAA 1680
Db 1621 GTCAAGCTGTCACTGAATTAACCTTGACCTGGACCCGAGTGAACAGCTGATGGTATACAA 1680
QY 1681 GGTGCGCGCTGATTTATACCTTGACACATGTTGAAATGCGAAATCTGAATCGGAGATC 1740
Db 1681 GGTGCGCGCTGATTTATACCTTGACACATGTTGAAATGCGAAATCTGAATCGGAGATC 1740
QY 1741 ACCGAAGGTGTCATCATGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1800
Db 1741 ACCGAAGGTGTCATCATGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1800
QY 1801 CAGAAATTTATGCTGACGCTGTACAGAGAGAGTGTATGATGAGAGAGAGAGATAC 1860
Db 1801 CAGAAATTTATGCTGACGCTGTACAGAGAGAGTGTATGATGAGAGAGAGATAC 1860
QY 1861 CTGAAGCTGACGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1920
Db 1861 CTGAAGCTGACGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1920
QY 1921 CTGAATGCTGCGTGAATTAACCTGCTCAACAGGATTAACAGTGAAGTGAAGTGAAGTGAAG 1980
Db 1921 CTGAATGCTGCGTGAATTAACCTGCTCAACAGGATTAACAGTGAAGTGAAGTGAAGTGAAG 1980
QY 1981 GCGGTGAAGATGACGCTGTATCCGCGTATTAATCTCCAGAGGGATCAATCAACACAGGA 2040
Db 1981 GCGGTGAAGATGACGCTGTATCCGCGTATTAATCTCCAGAGGGATCAATCAACACAGGA 2040
QY 2041 TATGTGATACCTGAGCGAATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 2091
Db 2041 TATGTGATACCTGAGCGAATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 2091

```

RESULT 3

```

PCT-US95-06994-4
; Sequence 4: Application PC/TUS9506994
; GENERAL INFORMATION:
; APPLICANT: Children's Hospital & Medical Center
; APPLICANT: University of Washington
; APPLICANT: Washington State University Research Foundation
; APPLICANT: TARR, PHILLIP I
; APPLICANT: BILGE, SIMA S
; APPLICANT: BESSER, THOMAS E
; APPLICANT: VARY JR, JAMES C
; TITLE OF INVENTION: ESCHERICHIA COLI O157:H7 EPITHELIAL ADHESIN
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSER: CHRISTENSEN, O'CONNOR, JOHNSON AND KINDNESSPLLC

```

```

; STREET: SUITE 2800, 1420 FIFTH AVENUE
; CITY: SEATTLE
; STATE: WASHINGTON
; COUNTRY: USA
; ZIP: WA 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06994
; FILING DATE: 07-JUN-95
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,714
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BRODERICK, THOMAS F
; REGISTRATION NUMBER: 31,332
; REFERENCE/DOCKET NUMBER: CHOR-18591
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 682 8100
; TELEFAX: (206) 224 0779
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2091 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; DESCRIPTION: corresponds to SEQ ID NO:1,
; DESCRIPTION: nucleotides 3036-5126
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli O157: H7
; STRAIN: 86-24 NALR
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..2088
; PCT-US95-06994-4

```

Query Match 100.0%; Score 2091; DB 5; Length 2091;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2091; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ATCGGAATAACCACTCTGGCTCCGTTAGTCAATTCCTGTCCTGGATTTTCAGCCAGCAGC 60
Db 1 ATCGGAATAACCACTCTGGCTCCGTTAGTCAATTCCTGTCCTGGATTTTCAGCCAGCAGC 60
QY 61 ATAGCTGCTGCAGAGATGTGATGTTCTCGGATCCGCTATGAGAAAGTGTACT 120
Db 61 ATAGCTGCTGCAGAGATGTGATGTTCTCGGATCCGCTATGAGAAAGTGTACT 120
QY 121 AACGACGCCCGCAGTGTTCCTGTGATTAGCCAGGAGAAATTCAGTCCAGCCAGTACCAC 180
Db 121 AACGACGCCCGCAGTGTTCCTGTGATTAGCCAGGAGAAATTCAGTCCAGCCAGTACCAC 180
QY 181 GATCTCGCGAGGCTCTGAGATCAGTACGAGGTGTGGATGTTGAAAGTGTACGGGTAAA 240
Db 181 GATCTCGCGAGGCTCTGAGATCAGTACGAGGTGTGGATGTTGAAAGTGTACGGGTAAA 240
QY 241 ACCGAGAGGCTGGAATCAGCATCCGAGGAATGCCAGGAAATCCAGCTGATCTGATT 300
Db 241 ACCGAGAGGCTGGAATCAGCATCCGAGGAATGCCAGGAAATCCAGCTGATCTGATT 300
QY 301 GATGTTGTTGTCAGGCGGAGAGAGTGAAGTGAATCCCAACGGTTTTCGCCAATGAAT 360
Db 301 GATGTTGTTGTCAGGCGGAGAGAGTGAAGTGAATCCCAACGGTTTTCGCCAATGAAT 360
QY 361 ACCGGGTTTCATGCCCTCTGCCCGCATTTGAGCTTATGAGGTTATCAGGGGGCCGATG 420

```

[illegible]

Qy	1501	TTCTCCAAAGAATGTGCTCTTATTCATATAATGATAACACCAATAGCTATGTAAACAGC	1560
Db	1501	TTCTCCAAAGAATGTGCTCTTATTCATATAATGATAACACCAATAGCTATGTAAACAGC	1560
Qy	1561	GGAAGCCCGGTGACCGGTGTGGAAATTTGCCGGACATTTGCCGTGTGGTCAGAGGAT	1620
Db	1561	GGAAGCCCGGTGACCGGTGTGGAAATTTGCCGGACATTTGCCGTGTGGTCAGAGGAT	1620
Qy	1621	GTCAACGTGTCACTGAANTTACACTGACCCGGAAGTGAAACAACGTGATGTGTGATAACAAA	1680
Db	1621	GTCAACGTGTCACTGAANTTACACTGACCCGGAAGTGAAACAACGTGATGTGTGATAACAAA	1680
Qy	1681	GGTGGCCCGCTGAGTTATACCCCTGAAACAATGGTGAATGCGAAACTGAACTGCGCAGATC	1740
Db	1681	GGTGGCCCGCTGAGTTATACCCCTGAAACAATGGTGAATGCGAAACTGAACTGCGCAGATC	1740
Qy	1741	ACCGAAGAGTGGCATCATGGCTGGGTGCGGTATTCGCGGGAAAAACACACGTTTCACC	1800
Db	1741	ACCGAAGAGTGGCATCATGGCTGGGTGCGGTATTCGCGGGAAAAACACACGTTTCACC	1800
Qy	1801	CAGAAATTATTTCGTCACTGAGCGGTGTACAGAAAGAAAGTGTATGATGAGAAGGAGAAATAC	1860
Db	1801	CAGAAATTATTTCGTCACTGAGCGGTGTACAGAAAGAAAGTGTATGATGAGAAGGAGAAATAC	1860
Qy	1861	CTGAAACCCCTGGACGGTGGTGATGCAAGGTCTGTCTGTGGAAGATGACGGATGCCCTGACG	1920
Db	1861	CTGAAACCCCTGGACGGTGGTGATGCAAGGTCTGTCTGTGGAAGATGACGGATGCCCTGACG	1920
Qy	1921	CTGAATCTCGGGTGAAATAACCTGCTCAAACAAGATTACAGTGACGTGACCCCTGTACAGT	1980
Db	1921	CTGAATCTCGGGTGAAATAACCTGCTCAAACAAGATTACAGTGACGTGACCCCTGTACAGT	1980
Qy	1981	GCCGGTAAGAGTACGCTGTATGCGGTGTATTCTTCCAGACGGGATCATCAACAACAGGA	2040
Db	1981	GCCGGTAAGAGTACGCTGTATGCGGTGTATTCTTCCAGACGGGATCATCAACAACAGGA	2040
Qy	2041	TATGTGATACCTGAGCGAAATTACTGGATGTGCGTGAACATCATGTTCTCTGA	2091
Db	2041	TATGTGATACCTGAGCGAAATTACTGGATGTGCGTGAACATCATGTTCTCTGA	2091

RESULT 4
US-08-765-081-1
; Sequence 1, Application US/08765081
; Patent No. 5798260
; GENERAL INFORMATION:
; APPLICANT: Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
; TITLE OF INVENTION: Escherichia Coli 0157:H7 Epithelial Adhesin
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen,
; STREET: 2800 Pacific First Center, 1420 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette-3.5 inch, 1.44mb storage
; COMPUTER: IBM PC compatible/Pentium
; OPERATING SYSTEM: MS-Windows 3.1
; SOFTWARE: Word for Windows-6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/765,081
; FILING DATE: March 26, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06994
; FILING DATE: June 7, 1995
; APPLICATION NUMBER: US 08/265,714
; FILING DATE: June 24, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Shelton, Dennis K.
; REGISTRATION NUMBER: 26,997

```
/ REFERENCE/DOCKET NUMBER: CHOR-1-10286
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 1-206-682-8100; 1-206-224-0718 (direct)
/ TELEFAX: 1-206-224-0779
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 8041 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
/ HYPOTHEetical: NO
/ ANTI-SENSE: NO
/ ORIGINAL SOURCE:
/ ORGANISM: Escherichia coli O157:H7
/ STRAIN: 86-24 NALR
/ IMMEDIATE SOURCE:
/ CLONE: pEAR
/ US-08-765-081-1

Query Match      100.0%; Score 2091; DB 1; Length 8041;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2091; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCGGAATACCACTCTGGCTCCGCTGAGTCAATTCCTGCTCGGATTTTCAGCCAGCAGC 60
DB 3036 ATCGGAATACCACTCTGGCTCCGCTGAGTCAATTCCTGCTCGGATTTTCAGCCAGCAGC 3095

QY 61 ATAGCTGCTCAGAGGATGTGATGTTGCTCGGCATCCGGTATGAGAAAAGCTGACT 120
DB 3096 ATAGCTGCTCAGAGGATGTGATGTTGCTCGGCATCCGGTATGAGAAAAGCTGACT 3155

QY 121 AACGACGCCGAGTCTGTTCTGTGATTAGCAGGAGGATTCAGTCCAGCCAGTACCAC 180
DB 3156 AACGACGCCGAGTCTGTTCTGTGATTAGCAGGAGGATTCAGTCCAGCCAGTACCAC 3215

QY 181 GATCTCGCGGAGGCTCTGAGATCAGTAGAGGGTGTGGATGTTGAAAGTGGTACCGGTAAA 240
DB 3216 GATCTCGCGGAGGCTCTGAGATCAGTAGAGGGTGTGGATGTTGAAAGTGGTACCGGTAAA 3275

QY 241 ACCGAGGGCTGGAATCAGATCCGAGGATCCGAGGATCCAGCCAGTACACGCTGATGATT 300
DB 3276 ACCGAGGGCTGGAATCAGATCCGAGGATCCGAGGATCCAGCCAGTACACGCTGATGATT 3335

QY 301 GATGTTGTTGCTCAGGCGCGAAGCAGTACGCTGACTCCCAACGGTTTTTCTGCCATGAT 360
DB 3336 GATGTTGTTGCTCAGGCGCGAAGCAGTACGCTGACTCCCAACGGTTTTTCTGCCATGAT 3395

QY 361 ACCGGGTTCAATGCCCTCTGGCGGCATGAGCGTATTGAGGTTATCAGGGGCGGATG 420
DB 3396 ACCGGGTTCAATGCCCTCTGGCGGCATGAGCGTATTGAGGTTATCAGGGGCGGATG 3455

QY 421 TCCACACTGTATGGCTCTGATCGATGCGGCGGTGTGGTGAATATCATACCAAGAAAGAT 480
DB 3456 TCCACACTGTATGGCTCTGATCGATGCGGCGGTGTGGTGAATATCATACCAAGAAAGAT 3515

QY 481 GCAGACAAATGGCTCTCTTCGCTCAATCAGGGCTGAATCTCAGGAAAGCAAAATGG 540
DB 3516 GCAGACAAATGGCTCTCTTCGCTCAATCAGGGCTGAATCTCAGGAAAGCAAAATGG 3575

QY 541 GGTAAACAGCAGCGAGTTAATTTCTGGACAGTGGTCCCTTTGTGGATGATTCTGTGACG 600
DB 3576 GGTAAACAGCAGCGAGTTAATTTCTGGACAGTGGTCCCTTTGTGGATGATTCTGTGACG 3635

QY 601 CTGACAGTACGGGTAGCAGACAAACAGCGTACAGGGTTCAATCGGTCACATCACTGACGAT 660
DB 3636 CTGACAGTACGGGTAGCAGACAAACAGCGTACAGGGTTCAATCGGTCACATCACTGACGAT 3695

QY 661 ACAGCAGGACCGGATTTCTTATCCAGGAGTACAGAAATTAATCTTGGTGCAGT 720
DB 3696 ACAGCAGGACCGGATTTCTTATCCAGGAGTACAGAAATTAATCTTGGTGCAGT 3755

QY 721 CTTGACTGGAAGCGCTCGGAGCAGGATGTGCTCTGCTTTGATATGATACCAACCGGCGAG 780
DB 3756 CTTGACTGGAAGCGCTCGGAGCAGGATGTGCTCTGCTTTGATATGATACCAACCGGCGAG 3815

QY 781 CGTTATGATAACCGGATGGGCAACTGGGAGTCTCAGCGGGGATATGACCGGACCTTG 840
DB 3816 CGTTATGATAACCGGATGGGCAACTGGGAGTCTCAGCGGGGATATGACCGGACCTTG 3875

QY 841 CGCTATGAGCGAAACAAAATTTTCAGTCTGCTATGATCATCTTACCTTCGGAAACATCG 900
DB 3876 CGCTATGAGCGAAACAAAATTTTCAGTCTGCTATGATCATCTTACCTTCGGAAACATCG 3935

QY 901 AAATCGTATCTGAACTGGAACGAGACAGAAATAAAGGTCGTGAGCTTTGACGAGTGA 960
DB 3936 AAATCGTATCTGAACTGGAACGAGACAGAAATAAAGGTCGTGAGCTTTGACGAGTGA 3995

QY 961 CTGAAGCGCAAAATGGGGGCTTTCGGTTCAGCCCGGGGAGCTTAAGAAATGCAACCTT 1020
DB 3996 CTGAAGCGCAAAATGGGGGCTTTCGGTTCAGCCCGGGGAGCTTAAGAAATGCAACCTT 4055

QY 1021 ATCTGGAATTCATTACTGCTTACCCCTCTGGGAGAAATCTCATCTGTTACGGTGGGGGC 1080
DB 4056 ATCTGGAATTCATTACTGCTTACCCCTCTGGGAGAAATCTCATCTGTTACGGTGGGGGC 4115

QY 1081 GAGTTTCAGAGCTCGTCCATGAAAGACGGAGTTGCTTCCAGCAGACAGGTGAAATCTTTC 1140
DB 4116 GAGTTTCAGAGCTCGTCCATGAAAGACGGAGTTGCTTCCAGCAGACAGGTGAAATCTTTC 4175

QY 1141 CGGCAGAAAGCTGTCGGTATTTGCTGAGGATGAGTGGCATCTCAGCGATGCACATTGCG 1200
DB 4176 CGGCAGAAAGCTGTCGGTATTTGCTGAGGATGAGTGGCATCTCAGCGATGCACATTGCG 4235

QY 1201 CTGACTCGCGGACGCGCTATGAAACATCATGAGCAATTCGGGGGACACTTCAGTCCGCT 1260
DB 4236 CTGACTCGCGGACGCGCTATGAAACATCATGAGCAATTCGGGGGACACTTCAGTCCGCT 4295

QY 1261 GCATATCTGCTCTGGATGTGGCAGATGCTGACGCTGAAAGCGGTGTGACCGGGA 1320
DB 4296 GCATATCTGCTCTGGATGTGGCAGATGCTGACGCTGAAAGCGGTGTGACCGGGA 4355

QY 1321 TATAAGCACCCAGAAATGGGGCAGCTACATAAAGGATTTAGTGTGTGTCGGGGCAGGA 1380
DB 4356 TATAAGCACCCAGAAATGGGGCAGCTACATAAAGGATTTAGTGTGTGTCGGGGCAGGA 4415

QY 1381 AAAACAAATCTACTTGTGTAACCCCGACCTGAAACCGGAGAGCGCTCAGTTATGAGCT 1440
DB 4416 AAAACAAATCTACTTGTGTAACCCCGACCTGAAACCGGAGAGAGCGCTCAGTTATGAGCT 4475

QY 1441 GGGGTGTATTACGATAACCCCGCGCTCTGAAATGCCAATGTACAGGTTTATGACTGAC 1500
DB 4476 GGGGTGTATTACGATAACCCCGCGCTCTGAAATGCCAATGTACAGGTTTATGACTGAC 4535

QY 1501 TTCTCCAAACAGATTTGCTCTTATTCATAAATGATAACCAATAGCTATGTAACACAG 1560
DB 4536 TTCTCCAAACAGATTTGCTCTTATTCATAAATGATAACCAATAGCTATGTAACACAG 4595

QY 1561 GGAAGGCCCGGTGACCGGTGGATTTGCGCGCACATTCGCGCTGTGCTGAGAGGAT 1620
DB 4596 GGAAGGCCCGGTGACCGGTGGATTTGCGCGCACATTCGCGCTGTGCTGAGAGGAT 4655

QY 1621 GTACGCTGTCACTGAATTTACCTGACCCGAGTGAACCAACAGCTGATGGTATAACAA 1680
DB 4656 GTACGCTGTCACTGAATTTACCTGACCCGAGTGAACCAACAGCTGATGGTATAACAA 4715

QY 1681 GGTGCGCGCTGAGTTATACCCCTGACCATGCTGAATCGGAACTGACCTGGCAGATC 1740
DB 4716 GGTGCGCGCTGAGTTATACCCCTGACCATGCTGAATCGGAACTGACCTGGCAGATC 4775

QY 1741 ACCGAAGAGTGGCATCATGGCTGGGTGCCGCTTATCGGGGAAAAACACCACTTTTACC 1800
DB 4776 ACCGAAGAGTGGCATCATGGCTGGGTGCCGCTTATCGGGGAAAAACACCACTTTTACC 4835

QY 1801 CAGAAATTTTCTGCTCAGTGGCGCTGTACAGAAAGATGTATGATGAGAAAGAGATATAC 1860
DB 1801 CAGAAATTTTCTGCTCAGTGGCGCTGTACAGAAAGATGTATGATGAGAAAGAGATATAC 1860
```


Db 4056 ATCTGTAATTCATTACTGCTTACCCCTCTGGAGAACTCATCTGTTACGGTGGGGGC 4115
 Qy 1081 GAGTTTCAGAGCTGTCATGAAGAAGAGAGAGTTGCTTCCAGCAGCAGTGAACCTTTC 1140
 Db 4116 GAGTTTCAGAGCTGTCATGAAGAAGAGAGAGTTGCTTCCAGCAGCAGTGAACCTTTC 4175
 Qy 1141 CGGAGAGAAAGCTGGTGGTATTTGCTGAGGATGAGTGGCATCTCACGGATGCACTTGCG 1200
 Db 4176 CGGAGAGAAAGCTGGTGGTATTTGCTGAGGATGAGTGGCATCTCACGGATGCACTTGCG 4235
 Qy 1201 CTGACTCGGGCAGCCGCTAATGAACATCATGAGCAATTCGGGGGACATTCAGTCCGCGT 1260
 Db 4236 CTGACTCGGGCAGCCGCTAATGAACATCATGAGCAATTCGGGGGACATTCAGTCCGCGT 4295
 Qy 1261 GCATATCTGGTCTGGGATGGCAGATGCCCTGGACGCTGAAGCGGTGACACGGGA 1320
 Db 4296 GCATATCTGGTCTGGGATGGCAGATGCCCTGGACGCTGAAGCGGTGACACGGGA 4355
 Qy 1321 TATAAGGCACCCAGAAATGGGSCAGCTACATAAAGGATAGTGGTGTGTCGGGCGAGGA 1380
 Db 4356 TATAAGGCACCCAGAAATGGGSCAGCTACATAAAGGATAGTGGTGTGTCGGGCGAGGA 4415
 Qy 1381 AAAACAATCTACTTGTGTAACCCGACCTGAAGCGGAGAGAGCGTCAAGTTATGAGGCT 1440
 Db 4416 AAAACAATCTACTTGTGTAACCCGACCTGAAGCGGAGAGAGCGTCAAGTTATGAGGCT 4475
 Qy 1441 GGGGTGTATTACGATAACCCGCGGCTCTGAATGCCAATGTACAGGTTTATGACTGAC 1500
 Db 4476 GGGGTGTATTACGATAACCCGCGGCTCTGAATGCCAATGTACAGGTTTATGACTGAC 4535
 Qy 1501 TTCTCCAAAGATGTCCTTATTCATTAATGAATGAACCAATAGCTATGTAACAGC 1560
 Db 4536 TTCTCCAAAGATGTCCTTATTCATTAATGAATGAACCAATAGCTATGTAACAGC 4595
 Qy 1561 GGAAGCGCGGTGACGCTGAGTGGTGAATTTCCGGGACATTCGCGGTGCTGAGAGGAT 1620
 Db 4596 GGAAGCGCGGTGACGCTGAGTGGTGAATTTCCGGGACATTCGCGGTGCTGAGAGGAT 4655
 Qy 1621 GTACAGCTGTACTGAATTAACCTGACCCGAGTGAACCAAGTGAACCAAGTGAATCAAA 1680
 Db 4656 GTACAGCTGTACTGAATTAACCTGACCCGAGTGAACCAAGTGAACCAAGTGAATCAAA 4715
 Qy 1681 GGTGCGCGCTGAGTATACCTGACACATGCTGAATCGGAACCTGGAACCTGCGAGATC 1740
 Db 4716 GGTGCGCGCTGAGTATACCTGACACATGCTGAATCGGAACCTGGAACCTGCGAGATC 4775
 Qy 1741 ACCGAAGAGGTGGCATCATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1800
 Db 4776 ACCGAAGAGGTGGCATCATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 4835
 Qy 1801 CAGAAATTTCTGTCACCTGAGGCTGTACAGAAAGTGTATCATGAGAAAGAGGATAC 1860
 Db 4836 CAGAAATTTCTGTCACCTGAGGCTGTACAGAAAGTGTATCATGAGAAAGAGGATAC 4895
 Qy 1861 CTGAAGCCTGGAGCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1920
 Db 4896 CTGAAGCCTGGAGCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 4955
 Qy 1921 CTGAATGCTGGTGGTGAATACCTGCTCAAAGAGTATGAGTGAAGTGAAGTGAAGTGAAGT 1980
 Db 4956 CTGAATGCTGGTGGTGAATACCTGCTCAAAGAGTATGAGTGAAGTGAAGTGAAGTGAAGT 5015
 Qy 1981 GCCGTAAGAGTACGCTGTATGCGGTGATTTACTTCCAGACGGGATCATCAACACAGGA 2040
 Db 5016 GCCGTAAGAGTACGCTGTATGCGGTGATTTACTTCCAGACGGGATCATCAACACAGGA 5075
 Qy 2041 TATGTGATACCTGAGCGAAATTAAGTGAATGCTGATGCTGATGCTGATGCTGATGCTGAT 2091
 Db 5076 TATGTGATACCTGAGCGAAATTAAGTGAATGCTGATGCTGATGCTGATGCTGATGCTGAT 5126

; Sequence 1, Application PC/TUS9506994
 ; GENERAL INFORMATION:
 ; APPLICANT: Children's Hospital & Medical Center
 ; APPLICANT: University of Washington
 ; APPLICANT: Washington State University Research Foundation
 ; APPLICANT: TARR, PHILLIP I
 ; APPLICANT: BILGE, SIMA S
 ; APPLICANT: BESSER, THOMAS E
 ; APPLICANT: VARY JR, JAMES C
 ; TITLE OF INVENTION: ESCHERICHIA COLI O157:H7 EPITHELIAL ADHESIN
 ; NUMBER OF SEQUENCES: 14
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: CHRISTENSEN, O'CONNOR, JOHNSON AND KINDNESSPLLC
 ; STREET: SUITE 2800, 1420 FIFTH AVENUE
 ; CITY: SEATTLE
 ; STATE: WASHINGTON
 ; COUNTRY: USA
 ; ZIP: WA 98101
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC Compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US95/06994
 ; FILING DATE: 07-JUN-95
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/265,714
 ; FILING DATE: 24-JUN-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: BRODERICK, THOMAS F
 ; REGISTRATION NUMBER: 31,332
 ; REFERENCE/DOCKET NUMBER: CHOR-18591
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 682 8100
 ; TELEFAX: (206) 224 0779
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 8041 base pairs
 ; TYPE: nucleic acid
 ; TOPOLOGY: single
 ; STRAINEDNESS: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; ORGANISM: Escherichia coli O157:H7
 ; STRAIN: 86-24 NALR
 ; IMMEDIATE SOURCE:
 ; CLONE: PEAR
 ; PCT-US95-06994-1

Query Match 100.0%; Score 2091; DB 5; Length 8041;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2091; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 ATCGGAATAACCACTCTGGCTTCGGTAGTCATTCCTCGGATTCCTCGGATTTTCAGCCAGCAGC 60
 Db 3036 ATCGGAATAACCACTCTGGCTTCGGTAGTCATTCCTCGGATTCCTCGGATTTTCAGCCAGCAGC 3095
 Qy 61 ATAGCTGCTGCAGAGGATGTGATGATTGTCTCGGCATCCGGCTATGAGAAAAAGCTGACT 120
 Db 3096 ATAGCTGCTGCAGAGGATGTGATGATTGTCTCGGCATCCGGCTATGAGAAAAAGCTGACT 3155
 Qy 121 AACGAGCCGCGCAGTGTTCGTGATTAGCCAGGAGGATTCGATCCAGCCAGTACAC 180
 Db 3156 AACGAGCCGCGCAGTGTTCGTGATTAGCCAGGAGGATTCGATCCAGCCAGTACAC 3215
 Qy 181 CATCTGGCGGAGGCTCTGAGATCAGTAGAGGGTGTGGATGTTGAAAGTGTACGGGTAAA 240
 Db 3216 CATCTGGCGGAGGCTCTGAGATCAGTAGAGGGTGTGGATGTTGAAAGTGTACGGGTAAA 3275

QY 241 ACCGAGGGCTGGAATCAGCATCCGAGGAATCCGAGCAGTTCACCGCTGATCTGATT 300
 Db 3276 ACCGAGGGCTGGAATCAGCATCCGAGGAATCCGAGCAGTTCACCGCTGATCTGATT 3335
 QY 301 GATGGTTCCTGACGGCGGAAGAGTGAAGTGAATCCCAACCGTTTCTGCGCATGAAT 360
 Db 3336 GATGGTTCCTGACGGCGGAAGAGTGAAGTGAATCCCAACCGTTTCTGCGCATGAAT 3395
 QY 361 ACCGGGTTTCATGCCCTCTGCGCGCATTTGAGCGTATTGAGGTTATCAGGGGGCGGATG 420
 Db 3396 ACCGGGTTTCATGCCCTCTGCGCGCATTTGAGCGTATTGAGGTTATCAGGGGGCGGATG 3455
 QY 421 TCACACTGATGGCTGATCGGATGGCGGCTGGTGGTGAATATCATTTACGAGAAAGAT 480
 Db 3456 TCACACTGATGGCTGATCGGATGGCGGCTGGTGGTGAATATCATTTACGAGAAAGAT 3515
 QY 481 CGAGCAAAATGGCTCTCTCCGCTCAATGACGCGCTGAATCTCGAGGAAGCAACAAATGG 540
 Db 3516 CGAGCAAAATGGCTCTCTCCGCTCAATGACGCGCTGAATCTCGAGGAAGCAACAAATGG 3575
 QY 541 GGTAAACAGCAGCGAGTTAAATTTCTGGAGCAGTGGTCCCTTTGTGGATGATTTCTCAGC 600
 Db 3576 GGTAAACAGCAGCGAGTTAAATTTCTGGAGCAGTGGTCCCTTTGTGGATGATTTCTCAGC 3635
 QY 601 CTGAGGTACGGGTAGCAACAAACGCTGAGGTTTCACTCGTACATCAGTGGCGAT 660
 Db 3636 CTGAGGTACGGGTAGCAACAAACGCTGAGGTTTCACTCGTACATCAGTGGCGAT 3695
 QY 661 ACAGCAGGACCGGATTTCTTTATCCACGAGTACAGAAATTAATCTTGGTCAAGT 720
 Db 3696 ACAGCAGGACCGGATTTCTTTATCCACGAGTACAGAAATTAATCTTGGTCAAGT 3755
 QY 721 CTGAGTGAAGCGTCGAGCAGGATGTCTCTGGTTTGAATATGATACACCGCGGAG 780
 Db 3756 CTGAGTGAAGCGTCGAGCAGGATGTCTCTGGTTTGAATATGATACACCGCGGAG 3815
 QY 781 CGTTATGATTAACCGGATCGGAACTGGGAGTCTGACGGGGGATATGACCGGACCTG 840
 Db 3816 CGTTATGATTAACCGGATCGGAACTGGGAGTCTGACGGGGGATATGACCGGACCTG 3875
 QY 841 CGCTATGAGCGAAACAAATTTTCACTGGCTATGATCATATTTTCACTTTCGGAACATGG 900
 Db 3876 CGCTATGAGCGAAACAAATTTTCACTGGCTATGATCATATTTTCACTTTCGGAACATGG 3935
 QY 901 AAATCGTATCTGAACGAGCAGAGCAAGAAATTAAGTCTGAGTCTGACGAGTGA 960
 Db 3936 AAATCGTATCTGAACGAGCAGAGCAAGAAATTAAGTCTGAGTCTGACGAGTGA 3995
 QY 961 CTGAAGCGGACAAATGGGGCTTGGCGGTGAGCGCGGAGCTTAAGGAATCGAACCTT 1020
 Db 3996 CTGAAGCGGACAAATGGGGCTTGGCGGTGAGCGCGGAGCTTAAGGAATCGAACCTT 4055
 QY 1021 ATCTGAATTCATCTGTTACCTCTGAGGAGAAATCTCATCTGGTTACGGTGGGGGC 1080
 Db 4056 ATCTGAATTCATCTGTTACCTCTGAGGAGAAATCTCATCTGGTTACGGTGGGGGC 4115
 QY 1081 GAGTTTCAGAGCTCGCTCAATGAAGAGGAGTTGCTTTCGAGCAGCAGTGAATCTTC 1140
 Db 4116 GAGTTTCAGAGCTCGCTCAATGAAGAGGAGTTGCTTTCGAGCAGCAGTGAATCTTC 4175
 QY 1141 CGGAGAAAGCTGGTGGTATTTGCTGAGGATGAGTGCATCTCAGGATGCATTCGG 1200
 Db 4176 CGGAGAAAGCTGGTGGTATTTGCTGAGGATGAGTGCATCTCAGGATGCATTCGG 4235
 QY 1201 CTGACTCGGGCAGCGCTATGAACATCATGAGCAATTCGGGGGACACTTCAGTCCGGT 1260
 Db 4236 CTGACTCGGGCAGCGCTATGAACATCATGAGCAATTCGGGGGACACTTCAGTCCGGT 4295
 QY 1261 GCATATCTGGTCTGGGATGAGCAGATGCTGAGCAGCTGAAGCGGCTGACCAACGGGA 1320
 Db 4296 GCATATCTGGTCTGGGATGAGCAGATGCTGAGCAGCTGAAGCGGCTGACCAACGGGA 4355
 QY 1321 TATAAGGCAACCCAGAAATGGGGCAGCTACATAAAGGGAATTAGTGGTGTGTCGGGGCAGGGA 1380

Db 4356 TATAAGGCAACCCAGAAATGGGGCAGCTACATAAAGGGAATTAGTGGTGTCTCGGGCAGGGA 4415
 QY 1381 AAACCAATCTACTTGTGAACCCCGACCTCAAGCCGGAAGAGAGCGTCAGTTATGAGGCT 1440
 Db 4416 AAACCAATCTACTTGTGAACCCCGACCTCAAGCCGGAAGAGAGCGTCAGTTATGAGGCT 4475
 QY 1441 GGGTGTATTACGATAACCCCGCGGTCTGAATGCCAATGTCAAGTTTTATGACTGAC 1500
 Db 4476 GGGTGTATTACGATAACCCCGCGGTCTGAATGCCAATGTCAAGTTTTATGACTGAC 4535
 QY 1501 TTCTCCAAACAGATGTCTCTTATCCATAAATGATAACCAACCAATAGCTATGTAACAGC 1560
 Db 4536 TTCTCCAAACAGATGTCTCTTATCCATAAATGATAACCAACCAATAGCTATGTAACAGC 4595
 QY 1561 GGAAGGCCCGGTGGACGGTGTGGAATTTCCCGGCACATTTCCGGCTGTGGTCAGAGGAT 1620
 Db 4596 GGAAGGCCCGGTGGACGGTGTGGAATTTCCCGGCACATTTCCGGCTGTGGTCAGAGGAT 4655
 QY 1621 GTCAAGCTGTCACTGAATTTACCTGACCCGGAAGTGAACCAACGTGATGTTGTAACAAA 1680
 Db 4656 GTCAAGCTGTCACTGAATTTACCTGACCCGGAAGTGAACCAACGTGATGTTGTAACAAA 4715
 QY 1681 GGTCCGCGCTGAGTTATATCCCTGAAACACATGGTGAATCGAAACTGAACTGGCAGATC 1740
 Db 4716 GGTCCGCGCTGAGTTATATCCCTGAAACACATGGTGAATCGAAACTGAACTGGCAGATC 4775
 QY 1741 ACCGAGAGGTGGCATCATGGTGGGTCGCCGTTCATCGCGGGAACCAACACGTTTCACC 1800
 Db 4776 ACCGAGAGGTGGCATCATGGTGGGTCGCCGTTCATCGCGGGAACCAACACGTTTCACC 4835
 QY 1801 CAGAAATTTCTGTCACATGAGCGTGTACAGAAAGTGTATGATGAGAAAGGAGATAC 1860
 Db 4836 CAGAAATTTCTGTCACATGAGCGTGTACAGAAAGTGTATGATGAGAAAGGAGATAC 4895
 QY 1861 CTGAAGCCTGGACGGTGGTGGATGCAAGTCTGCTGGAAGATGACGGAATGCCCTGACG 1920
 Db 4896 CTGAAGCCTGGACGGTGGTGGATGCAAGTCTGCTGGAAGATGACGGAATGCCCTGACG 4955
 QY 1921 CTGAATGCTCGGTTGAATTAACCTGCTCAACAGAGTTACAGTGAAGTGTGAGTGTGAGT 1980
 Db 4956 CTGAATGCTCGGTTGAATTAACCTGCTCAACAGAGTTACAGTGAAGTGTGAGTGTGAGT 5015
 QY 1981 GCCGGTGAAGTACGCTGTATGCGCGGTGATTTCTTCAGACGGGATCATCAACACAGGA 2040
 Db 5016 GCCGGTGAAGTACGCTGTATGCGCGGTGATTTCTTCAGACGGGATCATCAACACAGGA 5075
 QY 2041 TATGTGATCTGAGCGAAATTTACTGGAATGTGCTGAATCATCACTGATCTGATCTGTA 2091
 Db 5076 TATGTGATCTGAGCGAAATTTACTGGAATGTGCTGAATCATCACTGATCTGTA 5126

RESULT 7

US-09-453-702B-57/c
 ; Sequence 57, Application US/09453702B
 ; Patent No. 6365723
 ; GENERAL INFORMATION:
 ; APPLICANT: Blattner, Frederick R.
 ; Burland, Valerie
 ; Perna, Nicole T.
 ; Plunkett, Guy
 ; Welch, Rod
 ; TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
 ; NUMBER OF SEQUENCES: 265
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Quarles & Brady
 ; STREET: 1 South Pinckney Street
 ; CITY: Madison
 ; STATE: WI
 ; COUNTRY: US
 ; ZIP: 53701-2113
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage

COMPUTER: IBM PC compatible	49391	CTGCAGTACGCGGTAGCACACAAACGCGTCAAGGTTTCATCGGTACATCACTGAGCGAT	49333
OPERATING SYSTEM: PC-DOS/MS-DOS	661	ACAGCAGGACGCGGTATTCCTTATCCACGGAGTACAGAAATTAATAATCTTGGTCAAGT	720
SOFTWARE: Word Perfect 8.0	49331	ACAGCAGGACGCGGTATTCCTTATCCACGGAGTACAGAAATTAATAATCTTGGTCAAGT	49272
CURRENT APPLICATION NUMBER:	721	CTTGACCTGGAAGGCGTCGGAGCAGGATGTCTCTGTTTGTATATGATACACACCGGCGAG	780
APPLICATION NUMBER: US/09/453,702B	49271	CTTGACCTGGAAGGCGTCGGAGCAGGATGTCTCTGTTTGTATATGATACACACCGGCGAG	49212
FILING DATE: 03-Dec-1999	781	CGTTTATGATAACCGGGATGGGCAACTCGGGAGTCTGAACGGGGGATATGAACCGGACCTG	840
CLASSIFICATION: <Unknown>	49211	CGTTTATGATAACCGGGATGGGCAACTCGGGAGTCTGAACGGGGGATATGAACCGGACCTG	49152
PRIOR APPLICATION DATA:	841	CGCTATGAGCGAAACAAATTTTCAGCTGGCTATGATCATCTTTCACCTTCGGACATGG	900
APPLICATION NUMBER: 60/110,955	49151	CGCTATGAGCGAAACAAATTTTCAGCTGGCTATGATCATCTTTCACCTTCGGACATGG	49092
FILING DATE: 04-DEC-1998	901	AAATCGTATCTGAACCTGGAACGAGACAGAAATAAAGGTGCTGAGCTGTGTAGCGAGTGA	960
ATTORNEY/AGENT INFORMATION:	49091	AAATCGTATCTGAACCTGGAACGAGACAGAAATAAAGGTGCTGAGCTGTGTAGCGAGTGA	49032
NAME: Seay, Nicholas J.	961	CTGAAGCGCGACAAATGGGGGCTTTGCCGTCAGCCCGGGAGCTTTAAGAAATCGAACCTT	1020
REGISTRATION/DOCKET NUMBER: 27386	49031	CTGAAGCGCGACAAATGGGGGCTTTGCCGTCAGCCCGGGAGCTTTAAGAAATCGAACCTT	48972
REFERENCE/DOCKET NUMBER: 960296,95017	1021	ATCCCTGAAATTCATTACTGCTTACCCCTCTGGGAGAAATCTCATCTGTTACGGTGGGGGCG	1080
TELECOMMUNICATION INFORMATION:	48971	ATCCCTGAAATTCATTACTGCTTACCCCTCTGGGAGAAATCTCATCTGTTACGGTGGGGGCG	48912
TELEPHONE: (608) 251-5000	901	GAGTTTCAGAGCTCGTCCATGAAAGACGGAGTTGCTTGCACGACACAGGTTGAAACTTTC	48852
TELEFAX: (608) 251-9166	1141	CGGCAGAAAGCTGCTCGGTATTTGCTGAGGATGAGTGGCATCTCACGATGCACTTCGCG	1260
INFORMATION FOR SEQ ID NO: 57:	48851	CGGCAGAAAGCTGCTCGGTATTTGCTGAGGATGAGTGGCATCTCACGATGCACTTCGCG	48792
SEQUENCE CHARACTERISTICS:	1201	CTGACTCGCGGACGCGCTATGAAACATCATGAGCAATTCGGGGGACACTTCAGTCCGCGT	48732
LENGTH: 87563	48791	CTGACTCGCGGACGCGCTATGAAACATCATGAGCAATTCGGGGGACACTTCAGTCCGCGT	48732
TYPE: nucleic acid	1361	GCATATCTGCTCTGGGATGTGGCAGATGCTGGAACGCTGGAAGGCGGTGTGACACGGGA	1320
STRANDEDNESS: double	48731	GCATATCTGCTCTGGGATGTGGCAGATGCTGGAACGCTGGAAGGCGGTGTGACACGGGA	48672
TOPOLOGY: linear	1321	TATAAGGACCCACAGAAATGGGGAGCTTACATAAAGGATAGTGGTGTGTCCGGGACAGGA	1380
MOLECULE TYPE: DNA (genomic)	48671	TATAAGGACCCACAGAAATGGGGAGCTTACATAAAGGATAGTGGTGTGTCCGGGACAGGA	48612
SEQUENCE DESCRIPTION: SEQ ID NO: 57:	1381	AAAAACAAATCTACTTGGTAAACCCCGACCTGGAACGCGGAGAGAGCGGTACAGTATGAGCT	1440
US-09-453-702B-57	48611	AAAAACAAATCTACTTGGTAAACCCCGACCTGGAACGCGGAGAGAGCGGTACAGTATGAGCT	48552
Query Match 99.9%; Score 2089.4; DB 4; Length 87563;	1441	GGGGTGTATTACGATTAACCCCGCGCTGGAATGCCAATGTCCAGTCTTATGACTGAC	1500
Best Local Similarity 100.0%; Pred. No. 0;	48551	GGGGTGTATTACGATTAACCCCGCGCTGGAATGCCAATGTCCAGTCTTATGACTGAC	48492
Mismatches 0; Conservative 0; Gaps 0;	1501	TTCTCCCAACAGGATTTGCTCTTATTCCATAAATGATAACCAATAGCTATGTAACAGC	1560
	48491	TTCTCCCAACAGGATTTGCTCTTATTCCATAAATGATAACCAATAGCTATGTAACAGC	48432
	1561	GGAAAGCCCGGTGTGACCGGTGTGGAATTTGCGGCAATTTCCCGCTGTGGTCAAGGAT	1620
	48431	GGAAAGCCCGGTGTGACCGGTGTGGAATTTGCGGCAATTTCCCGCTGTGGTCAAGGAT	48372
	1621	GTCAAGCTGTCACTGAATTTACACCTCGAACCGAAGTGAACAAACGCTGATGGTGAACAAA	1680
	48371	GTCAAGCTGTCACTGAATTTACACCTCGAACCGAAGTGAACAAACGCTGATGGTGAACAAA	48312
	1681	GGTGGCGCGTGTAGTTATACCCCTGAAACATGTTGGTGAATGCGAAATCACTGAGGAGATC	1740
	48311	GGTGGCGCGTGTAGTTATACCCCTGAAACATGTTGGTGAATGCGAAATCACTGAGGAGATC	48252

QY 1489 TTTATGACTGCTTCCAAACAGATTGTCTTATTCCATAATGATAAC 1539
 Db 1492 TTTAATACAGATTAAAGATAGCTTACCAACTACGATACCGGTGAATC 1542

RESULT 9
 US-09-489-039A-962
 ; Sequence 962, Application US/09489039A
 ; Patent No. 6610836
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary Breton et. al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
 ; FILE REFERENCE: 2709.2004001
 ; CURRENT APPLICATION NUMBER: US/09/489,039A
 ; CURRENT FILING DATE: 2000-01-27
 ; PRIOR APPLICATION NUMBER: US 60/117,747
 ; PRIOR FILING DATE: 1999-01-29
 ; NUMBER OF SEQ ID NOS: 14342
 ; SEQ ID NO 962
 ; LENGTH: 2100
 ; TYPE: DNA
 ; ORGANISM: Klebsiella pneumoniae
 US-09-489-039A-962

Query Match 5.0%; Score 105.4; DB 4; Length 2100;
 Best Local Similarity 51.9%; Pred. No. 2.1e-24;
 Matches 238; Conservative 0; Mismatches 221; Indels 0; Gaps 0;
 QY 1060 CATCTGTTACGCTGGGGGGGAGTTTCAGAGTTCGTCATGAAAGACGAGTTGCTT 1119
 Db 1108 CAGCTGTTGACCTTCGGCGGATGCGCCACGACAACTGAAGTCCGCTCAACTG 1167
 QY 1120 GCGAGCAGGTGAACCTTCCGGCAGAAAGCTGCTGTTATTCGTAGGATGATGG 1179
 Db 1168 AGCAGCGCGGCGAGTCAACGCTGCGTACGAGTACGCCCTGTTTATCGAAGACGATG 1227
 QY 1180 CATCTCAGGATGACATTCGCTGAGTTCGCGGAGCGCTATCAACATCATAGCAATC 1239
 Db 1228 CGCATATCGAGCGCTTCGGCGTACGACCGGATTCGTATGACACCATCAGACTAT 1287
 QY 1240 GGGGACACTTCAGTCCGCTGATATCTGCTGAGATGTCGATGCTGAGCTGACGCTG 1299
 Db 1288 GCGCATCACTGGAGCCCGCGCTATCTGCTGTATTAACGCCACCGATACCGTCACTG 1347
 QY 1300 AAAGCGGTGTACCCAGCGATATAGGACCCAGATGGGCGAGCTACATAAGGGAT 1359
 Db 1348 AAAGCGGTGGGCGAGCGGCTTAAAGCCCGCTGCTGCTGAGCTTAAACCCGACTGG 1407
 QY 1360 AGTGGTGTGTCGGGCGAGGAAACAAATCTACTTGTAAACCCCGACTGAAACCGGAA 1419
 Db 1408 ACCACCACTCTGTCGCGCTCGTGACGATCGTGTAACCCGATCTGAAACCGGAA 1467
 QY 1420 GAGAGCTCAGTATGAGCTGGGTGTTATACGATTAACCCCGCTGTAATGCCAAT 1479
 Db 1468 ACCAGCGAAGCTTCGAGCTCGTCTCTACTACCGCGGGAAGAGGCTGGCTTGAAT 1527
 QY 1480 GTACAGTTTATGACTGACTTCTCCAAACAGATTGTC 1518
 Db 1528 GTCAAGGCGACATCACCACTTCCAGATTAATGTGAC 1566

RESULT 10
 US-08-480-510-1
 ; Sequence 1, Application US/08480510
 ; Patent No. 5747028
 ; GENERAL INFORMATION:
 ; APPLICANT: Calderwood, Stephen B.
 ; APPLICANT: Butters, Joan R.
 ; APPLICANT: Mekalanos, John J.
 ; TITLE OF INVENTION: HETEROLOGOUS ANTIGENS IN LIVE
 ; TITLE OF INVENTION: CELL
 ; TITLE OF INVENTION: VACCINE STRAINS

NUMBER OF SEQUENCES: 12
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Fish & Richardson
 STREET: 225 Franklin Street
 CITY: Boston
 STATE: Massachusetts
 COUNTRY: U.S.A.
 ZIP: 02110-2804
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
 COMPUTER: IBM PS/2 Model 50Z or 55SX
 OPERATING SYSTEM: MS-DOS (Version 5.0)
 SOFTWARE: WordPerfect (Version 5.1)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/480,510
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/08/020,501
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Clark, Paul T.
 REGISTRATION NUMBER: 30,162
 REFERENCE/DOCKET NUMBER: 00786/136001
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 542-5070
 TELEFAX: (617) 542-8906
 TELEX: 200154
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1535
 TYPE: nucleic acid
 STRANDEDNESS: both
 TOPOLOGY: linear
 US-08-480-510-1

Query Match 5.0%; Score 105; DB 1; Length 1535;
 Best Local Similarity 57.4%; Pred. No. 2.3e-24;
 Matches 214; Conservative 0; Mismatches 150; Indels 9; Gaps 1;
 QY 71 CAGAGATGATGATGATGCTCGGCATCCGCTATGAGAAAAGCTACTAACCCACCG 130
 Db 1168 CGGATGAACCATGCTGCTCACTCGCGGGGATACGCGCAAGTATTCAAAATGCAACG 1227
 QY 131 CCAGTGTTCGTGATTAGCCAGGAGAAATGCAAGTCCAGCCAGTACCAAGATCTGCGCG 190
 Db 1228 CCAGTATCAGTGTGATTTCAAGAGAAGATCTGGATCTCGCTATTACCGTATGACCG 1287
 QY 191 AGGCTCTGATCATGATGAGGCTGTGATGTTGAAAGTGTACGGGTAAACCCGAGGGC 250
 Db 1288 ATGCGCTAAAAGGCTACCGGCTGTGACAGTACCGGAGGGGGGCGATACACC----- 1340
 QY 251 TGGAAATCAGCATCCGAGGATGCGAGTACAGCTTACAGCTGATGATGATGCTGTC 310
 Db 1341 --GATATCAGCTTCGTTGATGGATCAACTACTCTTAICTTGTGATGATGATGATG 1398
 QY 311 GTCAGGGCGGAAGCAGTACGATGCTCCCAACCGTTTTTCTGCCATGAATACCGGGTCA 370
 Db 1399 GCCAAACCTCACGCGAGACCCGCTCCAAACAGCGATGTCGCGGGGCGATGAGCAAGTTGGT 1458
 QY 371 TGCCCCCTTCGCGCCATTCGAGGCTATGAGGTTATCAGGGGCGCGATGTCACACTGT 430
 Db 1459 TACCGCACCTCAAGCGGATTAACGATCGAGGTGATCCGTGGGCCCGATGTCACGCTGT 1518
 QY 431 ATGGCTCTGATGC 443
 Db 1519 ACGGCTCGATGC 1531

RESULT 11
 PCT-US94-01780-1
 ; Sequence 1, Application PC/TUS9401780
 ; GENERAL INFORMATION:

APPLICANT: Calderwood, Stephen B.
 APPLICANT: Butterton, Joan R.
 APPLICANT: Mekalanos, John J.
 TITLE OF INVENTION: HETEROLOGOUS ANTIGENS IN LIVE CELL
 TITLE OF INVENTION: VACCINE STRAINS
 NUMBER OF SEQUENCES: 12
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Fish & Richardson
 STREET: 225 Franklin Street
 CITY: Boston
 STATE: Massachusetts
 COUNTRY: U.S.A.
 ZIP: 02110-2804
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
 COMPUTER: IBM PS/2 Model 502 or 555X
 OPERATING SYSTEM: MS-DOS (Version 5.0)
 SOFTWARE: WordPerfect (Version 5.1)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US94/01780
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/020,501
 FILING DATE: February 22, 1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Clark, Paul T.
 REGISTRATION NUMBER: 30,162
 REFERENCE/DOCKET NUMBER: 00786/136001
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 542-5070
 TELEFAX: (617) 542-8906
 TELEX: 200154
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1535
 TYPE: nucleic acid
 STRANDEDNESS: both
 TOPOLOGY: linear
 PCT-US94-01780-1

Query Match 5.0%; Score 105; DB 5; Length 1535;
 Best Local Similarity 57.4%; Pred. No. 2.3e-24;
 Matches 214; Conservative 0; Mismatches 150; Indels 9; Gaps 1;
 QY 71 CAGAGGATGATGATGTTGTCGGCATCCGGTATGAGAAAAAGTGAATCAACGAGCG 130
 Db 1168 CGATGAACCATGGTGGTCACTCGCGGGGATACGCGCACTGATTCAAAATGCACAG 1227
 QY 131 CCAGTGTTCGTGATTACCGAGGAGGAATTCAGTCCAGCAGTACCAAGTCTGGCGG 190
 Db 1228 CCAGTATCAGTGTGATTTCAAGAGAGATCTGGAATCTCGCTATTACCGTGAATGACCG 1287
 QY 191 AGCTCTGAGATCAGTAGAGGGTGTGGATGTTGMAAGTGTACGGGTAAACCGGAGGCG 250
 Db 1288 ATGCGGTAAACCGTACCGGGTGTGACAGTACCGGAGGGGGCGATACCTACC----- 1340
 QY 251 TGAATATCAGATCCAGGAATGCCAGCAGTGTACAGCTGATGATGATGATGATGATG 310
 Db 1341 --GATATCAGATTCGTGGTATGGGATCAAACTATACCTTATCTTTGGTGGATGTAAGC 1398
 QY 311 GTCAGGGCGGAGCAGTACGATGACCTCCCAAGGTTTTCGCCATGATACCGGTTCA 370
 Db 1399 GCCAAACCTCAOCCAGAGCCCGTCCAAACAGCGATGGCCCGGCGATTTGCAAGGTTGT 1458
 QY 371 TGCCCTCTGGCGGCATTGACGTTATTCAGGTTATCAGGGGGCGGATGTCACACTGT 430
 Db 1459 TACCGCACTGCAAGCGATTGAACGATTCAGGTGATCCGTGGCCCGGATGTCCTACCTGT 1518
 QY 431 ATGGCTCTGATGC 443
 Db 1519 ACGGCTCGGATGC 1531

RESULT 12
 US-09-543-681A-3764
 ; Sequence 3764; Application US/09543681A
 ; Patent No. 6605709
 ; GENERAL INFORMATION:
 ; APPLICANT: GARY BRETON
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
 ; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 2709, 1002-001
 ; CURRENT APPLICATION NUMBER: US/09/543,681A
 ; CURRENT FILING DATE: 2000-04-05
 ; PRIOR APPLICATION NUMBER: US 60/128,706
 ; PRIOR FILING DATE: 1999-04-09
 ; NUMBER OF SEQ ID NOS: 8344
 ; SEQ ID NO 3764
 ; LENGTH: 2034
 ; TYPE: DNA
 ; ORGANISM: Proteus mirabilis
 ; US-09-543-681A-3764

Query Match 4.6%; Score 96.8; DB 4; Length 2034;
 Best Local Similarity 50.0%; Pred. No. 1.6e-21;
 Matches 276; Conservative 0; Mismatches 267; Indels 9; Gaps 1;
 QY 63 AGCTGTGCAGAGATGTGATGATTGTCTCGGCATCGGCTATCAGAAAAAGCTGACTAA 122
 Db 108 AGAAGATCCTGAAAGGTTACTGCTGACAAACCGCTTCGGGATTTAAACAGACTGTTGAAGA 167
 QY 123 CGAGCGCGCAGTGTCTGTGATTAGCCAGGAGGAATTCAGTCCAGCCAGTACCACGA 182
 Db 168 TGCTCCAGCTCTGTTTCAGTGTACTCGTGAACAATTAGAACAACAATCATCATCGTGA 227
 QY 183 TCTGGCGGAGCTCTGAGATCAGTAGAGGTGTGGATGTTGAAAGTGTGTCGGGTAAAC 242
 Db 228 TGTGACTGATGCGCTAAAGATGTTCCGGGGTGTCTAGT-----GACGGGGGGCGG 278
 QY 243 CGAGGGCTGGAATCAGCATCCGAGGAATGCCAGGATTCACGCTGATCTGATTGA 302
 Db 279 TAGCAGCTCTGATATTAGTATTCTGTTGATGATGATGATGATGATGATGATGATGATG 338
 QY 303 TGGTGTTCGTGAGGGCGGAAGCAGTGCAGTCCCAACCGTTCCTCCCATGAATAC 362
 Db 339 TGGTAAAGCTGTGCACTCTGTGAACCGCTCTAATAGTAACTCAGGTATTGAACA 398
 QY 363 CGGTTTCATGCCCTCTGGCCGATTCAGGCTATTCAGGTTATCAGGGGGCGGATGC 422
 Db 399 AGGTTGGTTTACCTCCTTTACCCGCGATTGAGCGTATTGAGGTTAGTAAAGAGCCCAATGTC 459
 QY 423 CACACTGTATGGCTCTCATGCGATGGCGGTGTGATGAATATCAATACCAAGAAAGATGC 482
 Db 459 TCTTTATATGGCTCTCATGCGATGGCGGTGTGATGAATATCAATATATTAAGCGTAAAGCGCA 518
 QY 483 AGACAAATGGCTCTCTTCCTCAATGAGGGCTGAATCTCGAGGAAGCAACAATGGGG 542
 Db 519 AAAAGATGGAACCTTATGTTTACGTGTGACACAAACGTTGACTGAGCGTAAAAATGAAGG 578
 QY 543 TAAACAGCAGCAGTTTAAATTTCTGGACAGTGGTCCCTTGTGATGATGATCTGTCAGCCT 602
 Db 579 AAATACAGGCAAGGTAGTTTCTATGCTGAGGTCATGATTGATGATGATGATGATGATGATG 638
 QY 603 GCAGGTACCGGG 614
 Db 639 AAAGCTGCAAGG 650

RESULT 13
 US-09-489-039A-703
 ; Sequence 703; Application US/09489039A
 ; Patent No. 6610836
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary Breton et. al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA

;; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS

;; FILE REFERENCE: 2709.2034001

;; CURRENT APPLICATION NUMBER: US/09/489,039A

;; CURRENT FILING DATE: 2000-01-27

;; PRIOR APPLICATION NUMBER: US 60/117,747

;; PRIOR FILING DATE: 1999-01-29

;; NUMBER OF SEQ ID NOS: 14342

;; SEQ ID NO 703

;; LENGTH: 2280

;; TYPE: DNA

;; ORGANISM: Klebsiella pneumoniae

US-09-489-039A-703

Query Match 3.3%; Score 68.8; DB 4; Length 2280;
Best Local Similarity 57.4%; Pred. No. 4.2e-12;
Matches 124; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 389 TTGAGCGTATTGAGGTTATCAGGGGCGCGATGCCACACTGTATGCTCTGTATGCGATGG 448

DB 524 TCGAACGGATTGAAGTTCTCCGCGGTCCGCGAGCGCGCTATGTTCCGGCGCGCGG 583

QY 449 GCGGTGTGGTGAATCATATACAGAAAGATGCAGACAAATGGCTCTTTCGTCATG 508

DB 584 GCGAGTGGTCAATATCATATACCAACGTCCTCGACCACTTGGCAGGTTGCTGTCTT 643

QY 509 CAGGCGTGAATCTGCAGGAAGCAACAAATGGGTAAACAGCAGCCAGTTTAAATTTCTGGA 568

DB 644 TCTTACCACACGCGGAAACAAACAAAGAGCAGCACCACCAATCGCGCTNACTTCAATC 703

QY 569 GCAGTGTCCCTTGTGGATGATCTGTACGCTGC 604

DB 704 TCAGCGGCCACTGGCGCGGCGAGGCGCTGACGATGC 739

RESULT 14

US-09-668-113A-1

;; Sequence 1, Application US/09668113A

;; Patent No. 6410703

;; GENERAL INFORMATION:

;; APPLICANT: Russo, Thomas A.

;; TITLE OF INVENTION: Identification of A Vaccine Candidate from an

;; FILE REFERENCE: 11520.0214

;; CURRENT APPLICATION NUMBER: US/09/668,113A

;; CURRENT FILING DATE: 2000-09-22

;; NUMBER OF SEQ ID NOS: 10

;; SEQ ID NO 1

;; LENGTH: 2495

;; TYPE: DNA

;; ORGANISM: Escherichia coli

;; FEATURE:

US-09-668-113A-1

Query Match 3.3%; Score 68.6; DB 4; Length 2495;

Best Local Similarity 55.1%; Pred. No. 5.2e-12;

Matches 134; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 376 CCTCTGCGCCCATTTGAGGTTATGAGGTTATCAGGGGCGCGATGCCACACTGTATGCC 435

DB 732 CCACCGGAAACAGGTTGAGCGTATTGAAGTATCCGCGCCCTCGCGCGCGCGCTACGGT 791

QY 436 TCTGATGCGATGGCGGCTGTGGTGAATATCATACCAGAAAGATGCAGACAAATGGCTC 495

DB 792 TCGGGGCGCGCGGGGGTGGTGAATCATATCCAAACGTTCCCAACCACTGGCAC 851

QY 496 TCTTCGCTCAATGCAGGCTGAAATCTGAGGAAAGCAAAATGGGGTAAACAGCAGCCAG 555

DB 852 GGTTCGCTCTCGTTATACACCAACAGCAGCGGAAAGTAGCGAAGAGGGCGCTACGGTCCG 911

QY 556 TTTAAATTTCTGGAGCAGTCCCTTGTGGATGATCTGTGATGCTGTGATGCTGTGATGCT 615

DB 912 GCCAATTCAGCTTATAGTGGGCTCTGGCTGTGGATGCTTCTTACCACCGGTTTGTAGT 971

QY 616 AGC 618

DB 972 AAC 974

RESULT 15

US-09-489-039A-6223

;; Sequence 6223, Application US/09489039A

;; Patent No. 6610836

;; GENERAL INFORMATION:

;; APPLICANT: Gary Bzeton et. al

;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA

;; PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS

;; FILE REFERENCE: 2709.2004001

;; CURRENT APPLICATION NUMBER: US/09/489,039A

;; CURRENT FILING DATE: 2000-01-27

;; PRIOR APPLICATION NUMBER: US 60/117,747

;; PRIOR FILING DATE: 1999-01-29

;; NUMBER OF SEQ ID NOS: 14342

;; SEQ ID NO 6223

;; LENGTH: 2388

;; TYPE: DNA

;; ORGANISM: Klebsiella pneumoniae

US-09-489-039A-6223

Query Match 3.0%; Score 63.4; DB 4; Length 2388;

Best Local Similarity 54.5%; Pred. No. 2.8e-10;

Matches 127; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 376 CCTCTGCGCCCATTTGAGGTTATGAGGTTATCAGGGGCGCGATGCCACACTGTATGCC 435

DB 583 CCGCGCGAGATGATCGAACGTTATCGAAGTGAATTCGGCGCGCGCGCGCTACGGC 642

QY 436 TCTGATGCGATGGCGGCTGTGGTGAATATCATACCAGAAAGATGCAGACAAATGGCTC 495

DB 543 AACGGCGCGCGCGCGGCTGTGGTGAATATCATACCAGAAAGATGCAGACAAATGGCTC 702

QY 496 TCTTCCGTCATATGAGGCTGAATCTTCAGGAAGCAAAATGGGGTAAACAGCAGCCAG 555

DB 703 GGCTCATGGAACACCTATATCAACGCGCGCGAGCAAGGATGAAGGCTCCACCAACGC 762

QY 556 TTTAAATTTCTGGAGCAGTGTCCCTTGTGGATGATCTGTGATGCTGTGATGCTGTGATGCT 608

DB 763 ACTAACTTCAGCTCAGCGCCCACTCGCGCGCGGATTTAGCTTCGCGCTGTT 815

Search completed: October 12, 2004, 15:58:47

Job time : 161 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 12, 2004, 11:56:38 ; Search time 81.9 Seconds
(without alignments)
10846.132 Million cell updates/sec

Title: US-10-625-972-4
Perfect score: 2091
Sequence: 1 atgcgaataaacactctggc.....cgctgaactatcagttctga 2091

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 212409041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq 29Jan04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002s:*
7: Geneseqn2003as:*
8: Geneseqn2003bs:*
9: Geneseqn2003cs:*
10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2091	100.0	8041	2	AAT10105
C 2	2089.4	99.9	87563	8	Adi19044 E. coli 0
C 3	2087.8	99.8	86248	9	Adc00087 Enterocae
C 4	2079.8	99.5	76804	6	AbS78942 E. coli C
5	149.4	7.1	1980	7	AcF67549 Photorhab
6	149.4	7.1	110000	7	AcF65383 Photorhab
7	149.4	7.1	110000	7	Continuation (2 of
8	130.6	6.2	10353	4	Aas46270 DNA encod
9	125.6	6.0	2046	4	AbA89269 Escherich
10	125.6	6.0	12943	4	AbA89255 Escherich
11	105	5.0	1535	2	Aaq1387 Vibrio ch
12	104	5.0	1530	2	Aaq26542 Truncated
13	69.4	3.3	3675	5	Aas80035 DNA encod
14	68.6	3.3	2495	4	Aad03054 ironNec ge
C 15	66.6	3.2	4039	3	Aaa15184 DNA encod
16	64.2	3.1	758	4	Abk43522 DNA encod
C 17	64.2	3.1	822	4	Aas27059 cDNA enco
C 18	64.2	3.1	822	9	AdB93237 Human CDN
C 19	64.2	3.1	10731	4	Aas46247 DNA encod
C 20	63.2	3.0	2727	5	Aas93059 DNA encod
21	63.2	3.0	2808	5	Aas88838 DNA encod
22	57.6	2.8	686	4	Aas27475 cDNA enco
23	57.6	2.8	686	4	Abk43834 DNA encod

24	57.6	2.8	686	9	ADB93653 Human CDN
25	56.4	2.7	2790	8	Ada29178 DNA encod
26	52.2	2.5	2000	7	Ada71938 Rice gene
27	51.8	2.5	296	6	Abn75479 Human gly
28	51.6	2.5	1845	3	Aaz38967 Escherich
C 29	51.4	2.5	856	6	Abq29448 Oligonuc1
30	51.4	2.5	856	6	Abq29449 Oligonuc1
31	50.8	2.4	687	6	Abq3380 Oligonuc1
C 32	50.8	2.4	687	6	Abq3381 Oligonuc1
C 33	50.4	2.4	687	6	Abq3378 Oligonuc1
C 34	50.4	2.4	687	6	Abq3379 Oligonuc1
C 35	50.2	2.4	856	6	Abq29447 Oligonuc1
36	50.2	2.4	856	6	Abq29446 Oligonuc1
37	49.8	2.4	247	6	Abn26822 Human ORF
38	48.6	2.3	1963	4	AbA89976 Escherich
C 39	48.6	2.3	12264	4	AbA89966 Escherich
40	48.6	2.3	23654	6	AbS78844 E. coli C
41	48.4	2.3	1844	3	Aaz38968 Salmonell
42	47.8	2.3	885	5	Aaf26439 Pseudomon
43	47	2.2	2544	2	Aax91730 Porphorym
44	47	2.2	2634	2	Aax91603 Porphorym
45	47	2.2	12677	4	Aas46236 DNA encod

ALIGNMENTS

RESULT 1
AAT10105
ID AAT10105 standard; DNA; 8041 BP.

XX AAT10105;

DT 13-MAY-1996 (first entry)

DE Adherence conferring plasmid pear.

KW adhesin; plasmid pear; vector; vaccine; intestine colonisation; ds.

OS Escherichia; coli.

OS Synthetic.

OS Chimeric.

XX Key

Location/Qualifiers

FT primer_bind

FT complement(2867..2890)

FT /tag= a

FT /note= "primer for E.coli O157:H7 DNA"

FT primer_bind

FT complement(2895..2914)

FT /tag= b

FT /note= "primer for pear"

FT CDS

FT 3036..5126

FT /tag= c

FT /product= "adhesin"

FT /trans_except= pos:3912..3914; aa:Gly

FT /note= "claim 1, page 36"

FT misc_feature

FT 3271..3310

FT /tag= d

FT /note= "region of transposon TnpHoA insertion"

FT misc_feature

FT 3801..3840

FT /tag= e

FT /note= "region of transposon TnpHoA insertion"

FT primer_bind

FT complement(5159..5183)

FT /tag= g

FT /note= "primer for E.coli O157H7 DNA"

FT primer_bind

FT complement(5176..5196)

FT /tag= f

FT /note= "primer for pear"

FT CDS

FT complement(6449..7024)

FT /tag= h

FT /note= "terE homologue"

FT CDS

FT complement(7092..7670)

FT /tag= i

FT /note= "terD homologue"

XX WO9600233-A1.
PN
XX
XX PD 04-JAN-1996.
XX PF 07-JUN-1995; 95WO-US006994.
XX PR 24-JUN-1994; 94US-00265714.
XX PA (CHIL-) CHILDREN'S HOSPITAL & MEDICAL CENT.
PA (UNIW) UNIV WASHINGTON.
PA (UNIW) UNIV WASHINGTON STATE RES FOUND.
XX
XX PI Tarr PI, Bilge SS, Besser TE, Vary JC;
XX
XX WPI; 1996-068826/07.
DR P-PSDB; AAR75366.
XX
XX Chromosomal DNA from E. coli O157:H7 encoding epithelial adhesin -
PT isolated on plasmid pSC (overlap), for use as a vaccine to mediate
PT bacterial colonisation of bovine intestine.
XX
XX Disclosure; Page 23-30; 42pp; English.
XX
XX Adherence conferring plasmid pear (AAT10105) comprises Escherichia coli
CC O157:H7 chromosomal DNA plus the Stratagene SK+ vector. It includes 3
CC open reading frames, 2 of which are homologues of tere and terD genes
CC necessary for tellurite resistance. The third ORF is homologous to the
CC irgA gene and encodes an adhesin (AAR75366) that enables E. coli O157:H7,
CC an antibiotic-resistant, virulent and common food-borne pathogen, to
CC adhere to epithelial cells. This ORF can be utilised in the prodn. of
CC adhesin for use as a vaccine to prevent disease or colonisation of
CC mucosal surfaces by O157:H7
XX
SQ Sequence 8041 BP; 2241 A; 1807 C; 1838 G; 2126 T; 0 U; 29 Other;

Query Match 100.0%; Score 2091; DB 2; Length 8041;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2091; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCGAATAACCACTCTGGCTTCGGTAGTCATTCCTCGTCTCGGATTTTCAGCCAGCAGC 60
DB 3036 ATGCGAATAACCACTCTGGCTTCGGTAGTCATTCCTCGTCTCGGATTTTCAGCCAGCAGC 3095

QY 61 ATAGCTCTCGAGGAGTGTGATGTTCTCGGCATCCGGCTATGAGAAAGCTGACT 120
DB 3096 ATAGCTCTCGAGGAGTGTGATGTTCTCGGCATCCGGCTATGAGAAAGCTGACT 3155

QY 121 AACGCAGCCGCGAGTGTTCCTGTGATTAGCCAGGAGGAAATTCAGTCCAGCCAGTACCAC 180
DB 3156 AACGCAGCCGCGAGTGTTCCTGTGATTAGCCAGGAGGAAATTCAGTCCAGCCAGTACCAC 3215

QY 181 GATCTGCGGAGGCTCTGAGATCAGTAGAGGGTGTGATTTGAAAGTGTGACGGTAA 240
DB 3216 GATCTGCGGAGGCTCTGAGATCAGTAGAGGGTGTGATTTGAAAGTGTGACGGTAA 3275

QY 241 ACCGAGGGCTGGAATACAGATCCAGGAGGAAATGCCAGCAGTGTACACGCTGACTGATT 300
DB 3276 ACCGAGGGCTGGAATACAGATCCAGGAGGAAATGCCAGCAGTGTACACGCTGACTGATT 3335

QY 301 GATGGTTCCTGTCAGGCGGAGCAGTACGTCGACTCCCAACGGTTTTTCTGCCATGAAT 360
DB 3336 GATGGTTCCTGTCAGGCGGAGCAGTACGTCGACTCCCAACGGTTTTTCTGCCATGAAT 3395

QY 361 ACCGGTTTCATGCCCTCTCGCCGCCATTCAGCGTATTTGAGTTATCAGGGGCGCGATG 420
DB 3396 ACCGGTTTCATGCCCTCTCGCCGCCATTCAGCGTATTTGAGTTATCAGGGGCGCGATG 3455

QY 421 TCACACTGTATGGCTCTGATCGGATGGCGGTGTGGTGAATATCATTTACCAAGAAGAT 480
DB 3456 TCACACTGTATGGCTCTGATCGGATGGCGGTGTGGTGAATATCATTTACCAAGAAGAT 3515

QY 481 GCAGACAAATGGCTCTCTCCGTCAATGACGGGCTGAATCTGCAGGAAGACAAATGG 540

DB 3516 GCAGACAAATGGCTCTCTCCGTCAATGCAGGGCTGAATCTGCAGGAAGCAACAAATGG 3575
QY 541 GGTAAACAGAGCCAGTTTAAATTTCTGGAGCAGTGGTCCCTTTGGATGATCTGTGACG 600
DB 3576 GGTAAACAGAGCCAGTTTAAATTTCTGGAGCAGTGGTCCCTTTGGATGATCTGTGACG 3635
QY 601 CTGCAGGTACCGGTAGCACACAAACAGCGTTCAGGGTTCATCGGTCACTCATCTGAGCGAT 660
DB 3636 CTGCAGGTACCGGTAGCACACAAACAGCGTTCAGGGTTCATCGGTCACTCATCTGAGCGAT 3695
QY 661 ACAGCAGCAGCGGTATTCCTTATCCCAACGAGTACACAGAAATTAATCTTTGGTGACGT 720
DB 3696 ACAGCAGCAGCGGTATTCCTTATCCCAACGAGTACACAGAAATTAATCTTTGGTGACGT 3755
QY 721 CTTGACTGGAAGGCGTGGAGCAGGATGTCTGTGGTTTGTATATGGATACCAACCCGCGAG 780
DB 3756 CTTGACTGGAAGGCGTGGAGCAGGATGTCTGTGGTTTGTATATGGATACCAACCCGCGAG 3815
QY 781 CGTTATGATAACCGGGATGGGCAACTGGGGAGTCTGACGGGGGATATGACCGGACCGCTG 840
DB 3816 CGTTATGATAACCGGGATGGGCAACTGGGGAGTCTGACGGGGGATATGACCGGACCGCTG 3875
QY 841 CGCTATGAGCGAACAATAATTTACGCTGGCTATGATCATATCTTTCACCTTCGGACATGG 900
DB 3876 CGCTATGAGCGAACAATAATTTACGCTGGCTATGATCATATCTTTCACCTTCGGACATGG 3935
QY 901 AAATCGTATCTGAACTGGAAACGAGACAGAAATAAAGTCTGTGAGCTTTGACGAGTGTA 960
DB 3936 AAATCGTATCTGAACTGGAAACGAGACAGAAATAAAGTCTGTGAGCTTTGACGAGTGTA 3995
QY 961 CTGAAGCCGCAAAATGGGGCTTGGCGGTGAGCGGGAGCTTAAGGAATTCGAACCTT 1020
DB 3996 CTGAAGCCGCAAAATGGGGCTTGGCGGTGAGCGGGAGCTTAAGGAATTCGAACCTT 4055
QY 1021 ATCTGGAATTCATTACTTCTTACCCCTCTGGGAGAAATCTCATCTGGTTACGTTGGGGGGC 1080
DB 4056 ATCTGGAATTCATTACTTCTTACCCCTCTGGGAGAAATCTCATCTGGTTACGTTGGGGGGC 4115
QY 1081 GAGTTTCAGAGCTCGTCCATGAAAGAGCGAGTTCTCTTGGCCAGACAGGTGAAACTTTC 1140
DB 4116 GAGTTTCAGAGCTCGTCCATGAAAGAGCGAGTTCTCTTGGCCAGACAGGTGAAACTTTC 4175
QY 1141 CGGCAGAAAGCTCGTCCGTATTTGCTGAGATCAGTGGCATCTCACGGATGCACTTTCG 1200
DB 4176 CGGCAGAAAGCTCGTCCGTATTTGCTGAGATGAGTGGCATCTCACGGATGCACTTTCG 4235
QY 1201 CTGACTGCGGCGCAGCCGCTATGAAACATCATGAGCAATTCGGGGGACACTTCAGTCCGCGT 1260
DB 4236 CTGACTGCGGCGCAGCCGCTATGAAACATCATGAGCAATTCGGGGGACACTTCAGTCCGCGT 4295
QY 1261 GCATATCTGGTCTGGATGTGGAGATGCCCTGGAGCTGAAAGCGGTGTGACCAACGGGA 1320
DB 4296 GCATATCTGGTCTGGATGTGGAGATGCCCTGGAGCTGAAAGCGGTGTGACCAACGGGA 4355
QY 1321 TATAAGGCAACCCAGAAATGGGCGAGCTACATAAAGGGATAGTGTGTCTCCGGGACGGGA 1380
DB 4356 TATAAGGCAACCCAGAAATGGGCGAGCTACATAAAGGGATAGTGTGTCTCCGGGACGGGA 4415
QY 1381 AAAACAAATCTACTTGTGTAAACCCGAGCTGAAAGCGGAGAGCGTCACTGTATGAGGCT 1440
DB 4416 AAAACAAATCTACTTGTGTAAACCCGAGCTGAAAGCGGAGAGAGCGTCACTGTATGAGGCT 4475
QY 1441 GGGGTGTATTACGATAACCCCGCGGTCTGAATGCCAATGTCAAGGTTTATGACTGAC 1500
DB 4476 GGGGTGTATTACGATAACCCCGCGGTCTGAATGCCAATGTCAAGGTTTATGACTGAC 4535
QY 1501 TTCTCCAAACAGATGTCTCTTATTCATATAATGATAACCAATAGCTATGTAACAGC 1560
DB 4536 TTCTCCAAACAGATGTCTCTTATTCATATAATGATAACCAATAGCTATGTAACAGC 4595
QY 1561 GGAAGGCCGGTGTGACGGTGTGGAATTTGCGGCACTTTCGGCTGTGGTTCAGAGGAT 1620

4596 GGAAGGCCCGGTGACGGTGGAAATTTGCCGCCACATTTGCCCTGTGTGACAGGAT 4655
 1621 GTACGGCTGTCACTGAATTAACCTGGACCCGGAAGTGAACACGTGATGGTGATAACAAA 1680
 4656 GTACGGCTGTCACTGAATTAACCTGGACCCGGAAGTGAACACGTGATGGTGATAACAAA 4715
 1681 GGTGGCCGCTGAGTTATACCTGACACATGGTGAATGGCAACTGAATGGCAGATC 1740
 4716 GGTGGCCGCTGAGTTATACCTGACACATGGTGAATGGCAACTGAATGGCAGATC 4775
 1741 ACCGAGAGTGGCTCATGGCTGGGTGCCCTTATCGCGGAAACACCCAGTTTCACC 1800
 4776 ACCGAGAGTGGCTCATGGCTGGGTGCCCTTATCGCGGAAACACCCAGTTTCACC 4835
 1801 CAGAATTTATCTCACTGAGCCCTGTACAGAAAGTGTATGATGAGAAAGGAGATAC 1860
 4836 CAGAATTTATCTCACTGAGCCCTGTACAGAAAGTGTATGATGAGAAAGGAGATAC 4895
 1861 CTGAAGCTGACCGGTGGTGGATGACAGTCTGTCTGGAAGTACACGATGCCCTGACG 1920
 4896 CTGAAGCTGACCGGTGGTGGATGACAGTCTGTCTGGAAGTACACGATGCCCTGACG 4955
 1921 CTGAATGTCGGTGAATTAACCTGCTCAACAGGATTAACAGTGAAGTGTACAGT 1980
 4956 CTGAATGTCGGTGAATTAACCTGCTCAACAGGATTAACAGTGAAGTGTACAGT 5015
 1981 GCGGTGAAGTACCGTGTATGCGGTGATTTCTCCAGACGGGATCATCAACACAGGA 2040
 5016 GCGGTGAAGTACCGTGTATGCGGTGATTTCTCCAGACGGGATCATCAACACAGGA 5075
 2041 TATGTGATACCTGAGCGAAATTAACCTGATGCTGCTGAACATACAGTTCTGA 2091
 5076 TATGTGATACCTGAGCGAAATTAACCTGATGCTGCTGAACATACAGTTCTGA 5126

RESULT 2
 ACD19044/c
 ID ACD19044 standard; DNA; 87563 BP.
 XX
 AC ACD19044;
 XX
 DT 27-OCT-2003 (revised)
 DT 21-AUG-2003 (first entry)
 XX
 DE E. coli 0157 unique DNA sequence OZID_57.
 XX
 KW OZID; ds; acute haemorrhagic colitis; haemolytic uraemic syndrome;
 KW food poisoning.
 XX
 OS Escherichia coli; strain 0157:H7.
 XX
 PN US2003023075-A1.
 XX
 PD 30-JAN-2003.
 XX
 PF 01-APR-2002; 2002US-00114170.
 XX
 PR 04-DEC-1998; 98US-0110955P.
 PR 03-DEC-1999; 99US-00453702.
 XX
 PR (BLAT/) BLATTNER F. R.
 PA (BURL/) BURLAND V. D.
 PA (PERN/) PERNA N. T.
 PA (PLUN/) PLUNKETT G.
 PA (WELC/) WELCH R.
 XX
 PI Blattner FR, Burland VD, Perna NT, Plunkett G, Welch R;
 XX
 DR WPI; 2003-479497/45.
 XX

New DNA sequences from Escherichia coli strain 0157:H7, useful for detecting E. coli 0157:H7 in a sample, or in designing diagnostic probes which can be used to distinguish strain 0157:H7 from strain K12 using

PF molecular techniques.
 XX Claim 16; SEQ ID NO 57; 33pp; English.
 XX
 CC The invention relates to an isolated DNA molecule comprising an E. coli strain 0157:H7 sequence selected from a clostridial cytotoxin-like gene, a urease gene cluster, a RTX toxin-like gene cluster, a locus of enterocyte effacement and 2 genes from its associated lymphocytic phage 933W (a putative serine/threonine kinase and a tail fibre gene). E. coli 0157:H7 can cause food poisoning, specifically acute haemorrhagic colitis (which can develop into haemolytic uraemic syndrome). Also included are an isolated DNA molecule comprising a nucleotide sequence identical to at least 25 contiguous nucleotides contained in DNA sequences selected from ACD18988-ACD19242 (being 255 E. coli 0157 DNA sequences which are not found in E. coli K12), a recombinant DNA construction comprising the DNA above and a method for detecting E. coli 0157:H7 (ATCC 43895) in a sample (or distinguishing between 0157 and K12) using a probe derived from one of the 255 sequences. The DNA sequences are useful in detecting E. coli 0157:H7 in a sample, for the early diagnosis of humans and livestock infected with 0157:H7, and in designing diagnostic probes which can be used to distinguish strain 0157:H7 from strain K12 using molecular techniques. The present sequence is one of the 255 E. coli 0157:H7 DNA sequence (termed OZID_1-OZID255). Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from the USPTO at seqdata.uspto.gov/sequence.html?docID=20030023075 (Updated on 27-OCT-2003 to standardise OS field)

XX
 SQ Sequence 87563 BP; 22620 A; 20384 C; 21612 G; 22935 T; 0 U; 12 Other;

Query Match 99.9%; Score 2089.4; DB 8; Length 87563;
 Best Local Similarity 100.0%; Pred. NC. 0;
 Matches 2090; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATCGGAATACCACTCTGGCTTCGGTGTCTCGTGTCTCCCTGTCTCGGATTTTCAGCCAGCAGC 60
 DB 49991 ATCGGAATACCACTCTGGCTTCGGTGTCTCCCTGTCTCGGATTTTCAGCCAGCAGC 49932

QY 61 ATAGCTCTGTCGAGAGGATGTGATGATGTCTCGGCATCCGGCTATGAGAAAGCTGACT 120
 DB 49931 ATAGCTCTGTCGAGAGGATGTGATGATGTCTCGGCATCCGGCTATGAGAAAGCTGACT 49872

QY 121 AACGACGCCAGCTGTTCTGTGATTAGCCAGGAGGAATTCAGTCAGCCAGCTACCCAC 180
 DB 49871 AACGACGCCAGCTGTTCTGTGATTAGCCAGGAGGAATTCAGTCAGCCAGCTACCCAC 49812

QY 181 GATCTGCGGAGGCTCTCAGATCAGTAGAGGGTGTGGATGTTGAAAGTGTACCGGTAAA 240
 DB 49811 GATCTGCGGAGGCTCTCAGATCAGTAGAGGGTGTGGATGTTGAAAGTGTACCGGTAAA 49752

QY 241 ACCGAGGGCTGGAATATCAGCATCCGAGGAATCCAGCCAGTTACAGCTGATCTGATT 300
 DB 49751 ACCGAGGGCTGGAATATCAGCATCCGAGGAATCCAGCCAGTTACAGCTGATCTGATT 49692

QY 301 GATGTTGTTCTGTCAGGCGGAAGCAGTGACGTGACTCCCAACGGTTTTTCTGCCATGAAT 360
 DB 49691 GATGTTGTTCTGTCAGGCGGAAGCAGTGACGTGACTCCCAACGGTTTTTCTGCCATGAAT 49632

QY 361 ACCGGTTTCATGCCCTCTGCGCCGCAATGAGCGTATGAGGTTATCAGGGGCCCATG 420
 DB 49631 ACCGGTTTCATGCCCTCTGCGCCGCAATGAGCGTATGAGGTTATCAGGGGCCCATG 49572

QY 421 TCCACATGATGCTCTGTGATGCGGTGGTGGTGAATATCATTTACCAGAAAGAT 480
 DB 49571 TCCACATGATGCTCTGTGATGCGGTGGTGGTGAATATCATTTACCAGAAAGAT 49512

QY 481 GCAGACAAATGGCTCTCTTCCGTCAATGCAGGGGTGAATCTGCAGGAAGCAAAATGG 540
 DB 49511 GCAGACAAATGGCTCTCTTCCGTCAATGCAGGGGTGAATCTGCAGGAAGCAAAATGG 49452

QY 541 GGTAAACAGCAGCCAGTTTAAATTTCTGGAGCAGTGGTCCCTTGTGGATGATTTCTGTGAGC 600
 DB 49451 GGTAAACAGCAGCCAGTTTAAATTTCTGGAGCAGTGGTCCCTTGTGGATGATTTCTGTGAGC 49392

QY 601 CTCGAGTACCGGTAGACACACACACAGCGTCAGGGTTTCATCGGTTCATCTACATCTGAGCGAT 660
 Db 49391 CTGAGGTACCGGTAGACACACACACAGCGTCAGGGTTTCATCGGTTCATCTACATCTGAGCGAT 49332
 QY 661 ACACGAGCGACCGGTATTCCTTATCCACGAGGTTCACAGAAATATATATCTTGTGTGACGCT 720
 Db 49331 ACACGAGCGACCGGTATTCCTTATCCACGAGGTTCACAGAAATATATATCTTGTGTGACGCT 49272
 QY 721 CTTGACTGGAAGCGGTGGAAGAGGATGCTCTGTTGATATGATACACACCGGAG 780
 Db 49271 CTTGACTGGAAGCGGTGGAAGAGGATGCTCTGTTGATATGATACACACCGGAG 49212
 QY 781 GCTTATGATACCGGGATGGCACTGGGAGTCTGACGGGGATATGACCGGACCGCTG 840
 Db 49211 GCTTATGATACCGGGATGGCACTGGGAGTCTGACGGGGATATGACCGGACCGCTG 49152
 QY 841 CGCTATGAGCGAAACAAAATTTGAGTGGCTATGATCATATCTTCACTTCGGAACATGG 900
 Db 49151 CGCTATGAGCGAAACAAAATTTGAGTGGCTATGATCATATCTTCACTTCGGAACATGG 49092
 QY 901 AATCGTATCTGAACTGGAAGAGACAGAAATAAAGTCTGAGCTTGTACGACGTGA 960
 Db 49091 AATCGTATCTGAACTGGAAGAGACAGAAATAAAGTCTGAGCTTGTACGACGTGA 49032
 QY 961 CTGAAGCGGACAAATGGGGCTTTGCCGCTCAGCCGGGAGCTTAAGGAATCGAACCTT 1020
 Db 49031 CTGAAGCGGACAAATGGGGCTTTGCCGCTCAGCCGGGAGCTTAAGGAATCGAACCTT 48972
 QY 1021 ATCTGTAAATTCATCTGCTTACCTCTCGGAGAAATCTCATCTGTTACGTTGGGGG 1080
 Db 48971 ATCTGTAAATTCATCTGCTTACCTCTCGGAGAAATCTCATCTGTTACGTTGGGGG 48912
 QY 1081 GAGTTTCAGAGCTCGTCCATGAAGACGAGGATGCTCTGCCAGACAGTGAATTTTC 1140
 Db 48911 GAGTTTCAGAGCTCGTCCATGAAGACGAGGATGCTCTGCCAGACAGTGAATTTTC 48852
 QY 1141 CGGAGAAAGCTGCTGCTGATTTGCTGAGGATGAGTGGCATCTCAGGATGACCTTGG 1200
 Db 48851 CGGAGAAAGCTGCTGCTGATTTGCTGAGGATGAGTGGCATCTCAGGATGACCTTGG 48792
 QY 1201 CTGACTCGGCGACCGCTATGAACATCATGAGCAATTCGGGGACACATTCAGTCCGCT 1260
 Db 48791 CTGACTCGGCGACCGCTATGAACATCATGAGCAATTCGGGGACACATTCAGTCCGCT 48732
 QY 1261 GCATATCTGCTGCTGAGTGGCAGATGCTCGACCGTGAAGCGCGGTGTGACCAACGGA 1320
 Db 48731 GCATATCTGCTGCTGAGTGGCAGATGCTCGACCGTGAAGCGCGGTGTGACCAACGGA 48672
 QY 1321 TATAGGCACCCAGATGGGGAGCTACATAAAGGATAGTGGTGTGTCGGGACAGGA 1380
 Db 48671 TATAGGCACCCAGATGGGGAGCTACATAAAGGATAGTGGTGTGTCGGGACAGGA 48612
 QY 1381 AAAACAAATCTACTTGTGTAACCCCGACCTGAAGCCGGAAGAGAGCGTCTAGTTATGAGCT 1440
 Db 48611 AAAACAAATCTACTTGTGTAACCCCGACCTGAAGCCGGAAGAGAGCGTCTAGTTATGAGCT 48552
 QY 1441 GGGGTGTTATGATATACCCCGCGCTGGAATGCCAATGTACAGGTTTTATGACTGAC 1500
 Db 48551 GGGGTGTTATGATATACCCCGCGCTGGAATGCCAATGTACAGGTTTTATGACTGAC 48492
 QY 1501 TTCTCAACAAAGATGCTCTTATTCATAAATGATAACCAATAGCTATGTAACACAG 1560
 Db 48491 TTCTCAACAAAGATGCTCTTATTCATAAATGATAACCAATAGCTATGTAACACAG 48432
 QY 1561 GGAAGCCCGGTGACCGGTGGAATTTGCCGACATATGCCGCTGTGTTGTCAGAGAT 1620
 Db 48431 GGAAGCCCGGTGACCGGTGGAATTTGCCGACATATGCCGCTGTGTTGTCAGAGAT 48372
 QY 1621 GTCAAGCTGTCTGATTAACCTGGACCGGAAGTGAACAAGTGTATGATTAACAA 1680
 Db 48371 GTCAAGCTGTCTGATTAACCTGGACCGGAAGTGAACAAGTGTATGATTAACAA 48312

QY 1681 GGTGCGCGCTGAGTTATACCCCTGAAACACATGTTGAATGCAAACTGAATGCGCAGATC 1740
 Db 48311 GGTGCGCGCTGAGTTATACCCCTGAAACACATGTTGAATGCAAACTGAATGCGCAGATC 48252
 QY 1741 ACCGAAGAGTGGCATCATGCTGGTGGTCCGTTTATCCGGGAAAAACACACGTTTACC 1800
 Db 48251 ACCGAAGAGTGGCATCATGCTGGTGGTGGTCCGTTTATCCGGGAAAAACACACGTTTACC 48192
 QY 1801 CAGAAATTTCTGCTACGAGCGCTACAGAGAAAGTGTATGATGAGAGGAGATAC 1860
 Db 48191 CAGAAATTTCTGCTACGAGCGCTACAGAGAAAGTGTATGATGAGAGGAGATAC 48132
 QY 1861 CTGAAGCTTGAAGCGGTGGTGGATGCTGCTGCTGGAAGATGACGGATGCCCTGACG 1920
 Db 48131 CTGAAGCTTGAAGCGGTGGTGGATGCTGCTGCTGGAAGATGACGGATGCCCTGACG 48072
 QY 1921 CTGAATGCTGGGTGAATAACCTGCTCAACAGGATTTACAGTGACGTGAGCTGTACAGT 1980
 Db 48071 CTGAATGCTGGGTGAATAACCTGCTCAACAGGATTTACAGTGACGTGAGCTGTACAGT 48012
 QY 1981 GCCGTGAAGTACGCTGTATGCCGCTGATTTCTCCAGACGGGATCATCAACAAACAGGA 2040
 Db 48011 GCCGTGAAGTACGCTGTATGCCGCTGATTTCTCCAGACGGGATCATCAACAAACAGGA 47952
 QY 2041 TATGTGATCTGAGCGGAAATTAATCTGATGTCGCTGAACTATCAAGTTCTGA 2091
 Db 47951 TATGTGATCTGAGCGGAAATTAATCTGATGTCGCTGAACTATCAAGTTCTGA 47901

RESULT 3
 ADC00087/c
 ID ADC00087 standard; DNA; 86248 BP.
 XX
 AC ADC00087;
 XX
 DT 04-DEC-2003 (first entry)
 XX
 DE Enterohaemorrhagic E. coli O157:H7-specific nucleic acid SEQ ID NO: 132.
 XX
 KW ds; gene; enterohaemorrhagic; anti-bacterial.
 XX
 OS Escherichia coli; O157:H7.
 XX
 FN JP2002355074-A.
 XX
 PD 10-DEC-2002.
 XX
 PF 24-JAN-2002; 2002JP-00015959.
 XX
 PR 24-JAN-2001; 2001JP-00112010.
 XX
 PA (UYTS-) UNIV TSUKUBA.
 XX
 DR WPI; 2003-451640/43.
 XX
 PT Enterohaemorrhagic Escherichia coli O157:H7-specific nucleic acid molecule
 and a polypeptide and its use, a polypeptide, a vector and a host cell.
 XX
 PS Claim 2; SEQ ID NO 132; 2067pp; Japanese.
 XX
 CC The invention relates to a novel enterohaemorrhagic Escherichia coli
 O157:H7-specific nucleic acid molecule. A polynucleotide of the invention
 has anti-bacterial activity. The polypeptide can be used in detection
 and/or treatment of O157:H7 infection. The nucleotide sequence of the
 genome of Enterohaemorrhagic E. coli O157:H7 was determined. The present
 invention represents an E. coli O157:H7-specific nucleic acid of the
 invention.
 XX
 SQ Sequence 86248 BP; 22338 A; 20003 C; 21303 G; 22604 T; 0 U; 0 Other;
 Query Match 99.8%; Score 2087.8; DB 9; Length 86248;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 2089; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATCGGAATAACCACTCTGGCTCCCGTAGTCAATCCCTGCTCGGATTTTCAGCCAGCAGC 60
 Db 50004 ATGCGAATAACCACTCTGGCTCCCGTAGTCAATCCCTGCTCGGATTTTCAGCCAGCAGC 49945
 QY 61 ATAGCTGCTGCAGAGATGATGATGCTCGGCATCGGCTATGAGAAAGCTGACT 120
 Db 49944 ATAGCTGCTGCAGAGATGATGATGCTCGGCATCGGCTATGAGAAAGCTGACT 49885
 QY 121 AACCGAGCCGCGAGTGTCTGTGATTAGCCAGGAGGAATTCGAGTCCAGCCAGTACCAC 180
 Db 49884 AACCGAGCCGCGAGTGTCTGTGATTAGCCAGGAGGAATTCGAGTCCAGCCAGTACCAC 49825
 QY 181 GATCTGGCGGAGGCTCTGAGATCAGTAGAGGTGTGATGTTGAAAGTGTTGACGGGTAAA 240
 Db 49824 GATCTGGCGGAGGCTCTGAGATCAGTAGAGGTGTGATGTTGAAAGTGTTGACGGGTAAA 49765
 QY 241 ACCGGAGGCTGGAAATCAGCATCCGAGGAATCCAGCCAGTTCACGCTGATCTGATT 300
 Db 49764 ACCGGAGGCTGGAAATCAGCATCCGAGGAATCCAGCCAGTTCACGCTGATCTGATT 49705
 QY 301 GATGTTGTTTGGTACGGCGGGAAGCAGTGAAGTGAATCCCAACGGTTTTCTGCCATGAAT 360
 Db 49704 GATGTTGTTTGGTACGGCGGGAAGCAGTGAAGTGAATCCCAACGGTTTTCTGCCATGAAT 49645
 QY 361 ACCGGTTTCAATGCCCCCTCTGGCCGCAATGAGCGTATGAGGTTATCAGGGGGCGGATG 420
 Db 49644 ACCGGTTTCAATGCCCCCTCTGGCCGCAATGAGCGTATGAGGTTATCAGGGGGCGGATG 49585
 QY 421 TCACACTGTATGGCTCTGATGCGATGGCGGTGTGTTGAATATCAATACCAAGAAAT 480
 Db 49584 TCACACTGTATGGCTCTGATGCGATGGCGGTGTGTTGAATATCAATACCAAGAAAT 49525
 QY 481 GCGACAAATGGCTCTCTCCGTCATATGACGGCTGAATCTGCGAAGAAACAAATGG 540
 Db 49524 GCGACAAATGGCTCTCTCCGTCATATGACGGCTGAATCTGCGAAGAAACAAATGG 49465
 QY 541 GGTAAACAGCAGCAGTAAATTTCTGAGCAGTGGTCCCTTGTGGATGATTTCTGTCAGC 600
 Db 49464 GGTAAACAGCAGCAGTAAATTTCTGAGCAGTGGTCCCTTGTGGATGATTTCTGTCAGC 49405
 QY 601 CTCAGGTACGCGGTAGCACAACAACAGCGTACGGTTCATCGGTCACTACGTAGCGAT 660
 Db 49404 CTCAGGTACGCGGTAGCACAACAACAGCGTACGGTTCATCGGTCACTACGTAGCGAT 49345
 QY 661 ACAGCAGCAGCGGTATCTCTATCCAGCGGTACAGAAATTAATATCTGTGTCAGCT 720
 Db 49344 ACAGCAGCAGCGGTATCTCTATCCAGCGGTACAGAAATTAATATCTGTGTCAGCT 49285
 QY 721 CTTTGAATGGAAGCGTCGGAGCAGGATGTCTCTGGTTTGATATGGATACCCCGGCAG 780
 Db 49284 CTTTGAATGGAAGCGTCGGAGCAGGATGTCTCTGGTTTGATATGGATACCCCGGCAG 49225
 QY 781 CGTTATGATTAACCGGATGGGCAACTGGGAGTCTGACGGGGGATATGACGGACCTG 840
 Db 49224 CGTTATGATTAACCGGATGGGCAACTGGGAGTCTGACGGGGGATATGACGGACCTG 49165
 QY 841 CGCTATGAGCGAAACAAATTTTTCAGCTGGCTATGATCATACTTTTCACTTTGGAAACATGG 900
 Db 49164 CGCTATGAGCGAAACAAATTTTTCAGCTGGCTATGATCATACTTTTCACTTTGGAAACATGG 49105
 QY 901 AAATCGTATCTGAATCTGGAACGAGACAGAAATTAAGGTGCTGAGCTTGTAGCGAGTGA 960
 Db 49104 AAATCGTATCTGAATCTGGAACGAGACAGAAATTAAGGTGCTGAGCTTGTAGCGAGTGA 49045
 QY 961 CTGAGCGCAAAATGGGGGCTTGGCGTACCGCGGAGCTTAAGGAATCGAACCTT 1020
 Db 49044 CTGAGCGCAAAATGGGGGCTTGGCGTACCGCGGAGCTTAAGGAATCGAACCTT 48985
 QY 1021 ATCCTGAATTCATTACTGTTTACCCCTCTGGGAGAAATCTCATCTGGTTACGGTGGGGGC 1080
 Db 48984 ATCCTGAATTCATTACTGTTTACCCCTCTGGGAGAAATCTCATCTGGTTACGGTGGGGGC 48925

RESULT 4
 ABS78942/c
 ID ABS78942 standard; DNA; 76804 BP.
 XX

QY 1081 GAGTTTCAGAGCTCGTCCATGAAGACAGGAGTTGCTCTTCCAGCAGCAGGTGAACATTC 1140
 Db 48924 GAGTTTCAGAGCTCGTCCATGAAGACAGGAGTTGCTCTTCCAGCAGCAGGTGAACATTC 48865
 QY 1141 CGGCAGAAAAAGCTCGTCCGATTTTGTCTGAGGATGAGTGGCATCTCACGGATGCACTTGGC 1200
 Db 48864 CGGCAGAAAAAGCTCGTCCGATTTTGTCTGAGGATGAGTGGCATCTCACGGATGCACTTGGC 48805
 QY 1201 CTGACTCGGGGACCGCCTGATGAACATCATGAGCAATTCGGGGGACATTCAGTCCCGGT 1260
 Db 48804 CTGACTCGGGGACCGCCTGATGAACATCATGAGCAATTCGGGGGACATTCAGTCCCGGT 48745
 QY 1261 GCATATCTGTCTGGGATGTGGCAGATGCTTGAACGCTGAAAGCGCGTGTGACCAACGGGA 1320
 Db 48744 GCATATCTGTCTGGGATGTGGCAGATGCTTGAACGCTGAAAGCGCGTGTGACCAACGGGA 48685
 QY 1321 TATAAGCACCCAGAAATGGGGCAGCTACATAAAGGATTAAGTGTGTCTCGGSCAGGGA 1380
 Db 48684 TATAAGCACCCAGAAATGGGGCAGCTACATAAAGGATTAAGTGTGTCTCGGSCAGGGA 48625
 QY 1381 AAAACAAATCTACTTGTAAACCCGACCTGAAGCCGGAAGAGAGCGTCAGTTATGAGGCT 1440
 Db 48624 AAAACAAATCTACTTGTAAACCCGACCTGAAGCCGGAAGAGAGCGTCAGTTATGAGGCT 48565
 QY 1441 GGGGTGTATTACGATAAACCCCGCGCTCTGAATGCCAATGTCAAGGTTTATGACTGAC 1500
 Db 48564 GGGGTGTATTACGATAAACCCCGCGCGTCTAAA TGCCAAATGTCAAGGTTTATGACTGAC 48505
 QY 1501 TTCTCCAAACAGATTTCTTATTCATTAATGATAACCAACCANTAGCTATGTAACAGC 1560
 Db 48504 TTCTCCAAACAGATTTCTTATTCATTAATGATAACCAACCANTAGCTATGTAACAGC 48445
 QY 1561 GGAAGGCGCGGTGTCAGCGTGTGGAATTTGCCCGCACATTTGCCGCTGTGTGTGAGAGAT 1620
 Db 48444 GGAAGGCGCGGTGTCAGCGTGTGGAATTTGCCCGCACATTTGCCGCTGTGTGTGAGAGAT 48385
 QY 1621 GTCAAGCTGTCACTGAATTTACACTGGACCCGAAAGTGAACAACTGATGTTGTAACAAA 1680
 Db 48384 GTCAAGCTGTCACTGAATTTACACTGGACCCGAAAGTGAACAACTGATGTTGTAACAAA 48325
 QY 1681 GGTGCGCGCTGAGTTATACCCCTGAACACATGTTGAATGCGAACTGAACCTGGCAGATC 1740
 Db 48324 GGTGCGCGCTGAGTTATACCCCTGAACACATGTTGAATGCGAACTGAACCTGGCAGATC 48265
 QY 1741 ACCGAAGAGGTGGCATCATGCTGGTCCCGTTTATCGCGGAAAAACACCACTGTTCAAC 1800
 Db 48264 ACCGAAGAGGTGGCATCATGCTGGTCCCGTTTATCGCGGAAAAACACCACTGTTCAAC 48205
 QY 1801 CAGAAATTTCTGCTCACTGAGCGCTGTACAGAAAGTGTATGATGAGAAAGGAGATAC 1860
 Db 48204 CAGAAATTTCTGCTCACTGAGCGCTGTACAGAAAGTGTATGATGAGAAAGGAGATAC 48145
 QY 1861 CTGAAAGCTGGAAGCGGTGGTGGATGCGAGTCTGTCTGGAAGATGACGGATGCCCTGACG 1920
 Db 48144 CTGAAAGCTGGAAGCGGTGGTGGATGCGAGTCTGTCTGGAAGATGACGGATGCCCTGACG 48085
 QY 1921 CTGAATGCTGCGGTGAATTAACCTGCTCAA CAGGAATTAAGTGTGAGCTGAGCTGTACAGT 1980
 Db 48084 CTGAATGCTGCGGTGAATTAACCTGCTCAA CAGGAATTAAGTGTGAGCTGAGCTGTACAGT 48025
 QY 1981 GCCGTAGAGTACGCTGTATGCGGCTGATTACTTCCAGAGCGGATCATCAACACAGGA 2040
 Db 48024 GCCGTAGAGTACGCTGTATGCGGCTGATTACTTCCAGAGCGGATCATCAACACAGGA 47965
 QY 2041 TATGTGATACCTGAGCGAAATTAAGTGTGCTGCTGAACTATCATGTTCTGA 2091
 Db 47964 TATGTGATACCTGAGCGAAATTAAGTGTGCTGCTGAACTATCATGTTCTGA 47914

RESULT 6
ACF65383_0
WP Sequence split into 5 fragments LOCUS ACF65383 Accession ACF65383
WP Fragment Name Begin End
ACF65383_0 1 110000
ACF65383_1 10001 210000
ACF65383_2 20001 310000
ACF65383_3 30001 410000
ACF65383_4 40001 460203
ID ACF65383 standard; DNA; 460203 BP.
AC ACF65383;
XX.
DT 20-NOV-2003 (first entry)
XX
DE Photorhabdus luminescens nucleotide sequence #36.
XX
KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
KW detection; food; gene expression; plant; animal; microorganism; toxin;
KW antibiotic; biopesticide; virulence factor; disease model; plague;
KW whooping cough; gene; ds.
XX
OS Photorhabdus luminescens.
XX
XX WO200294867-A2.
XX
XX 28-NOV-2002.
XX
XX 07-FEB-2002; 2002WO-IB003040.
XX
XX 07-FEB-2001; 2001FR-00001659.
XX
XX (INSP) INST PASTEUR.
XX (CNRS) CNRS CENT NAT RECH SCI.
XX
XX Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;
XX Buchrieser C;
XX WPI; 2003-148459/14.
XX
XX The invention relates to the isolation of genes and their encoded
XX proteins from Photorhabdus luminescens. The isolated sequences are
XX sources of probes and primers for detecting the genome of P. luminescens
XX and related species; to study polymorphisms; for gene analysis and for
XX detection/amplification of the genes. Antibodies (Ab) raised against the
XX polypeptides encoded by the genes are used for detection/identification
XX of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
XX carry a gene-containing vector are used to select compounds that
XX modulate, regulate, induce or inhibit expression of the genes in plants,
XX animals or microorganisms other than P. luminescens and are able to alter
XX response or sensitivity to toxins and antibiotics produced by P.
XX luminescens. Cells transformed to express the genes are useful for
XX recombinant production of the proteins, particularly toxins and
XX antibacterial agents useful as insecticides, bactericides and fungicides. The
XX genes, proteins, vectors containing the genes and Ab are also useful
XX therapeutically (to treat microbial infection by bacteria or fungi that
XX are sensitive to P. luminescens-encoded toxins or antibiotics) and as
XX biopesticides. Other uses of the genes and the proteins are as virulence
XX factors and for identifying targets of human diseases for which P.
XX luminescens is a model (particularly plague and whooping cough). This
XX sequence represents one of the isolated P. luminescens genes
XX
XX Sequence 460203 BP; 128022 A; 88971 C; 107007 G; 136199 T; 0 U; 4 Other;
Query Match 7.1%; Score 149.4; DB 7; Length 110000;
Best Local Similarity 57.0%; Pred. No. 1.2e-35;

Matches 298; Conservative 0; Mismatches 216; Indels 9; Gaps 1;
QY 92 CGGCATCCGGCTATGAGAAAAGCTGACTAAACGAGCGCGCAGTGTCTTCTGATTAGCC 151
DB 58493 CGGCAGCAGGTTTTCACAGAGAAGATTGAAGATGCGCCAGCGCTATTCTCTGTGTAGCC 58552
QY 152 AGGAGGAATTTCAGTCCAGCCAGTACACAGATCTGGCGAGGCTCTGAGATCAGTAGAGG 211
DB 58553 GTGAGCAATTGGAAACTAAGCGTTATCTGTATGTGACTGTGCTGAAAGATGTACCGG 58612
QY 212 GTGTGGATGTTGAAAGTGGTACGGGTAAACCGGAGGCTGGAATACAGCATCCGAGGAA 271
DB 58613 GTGTGTTTACAGGTGGTGTCTAGCA-----GCAGTGATATCAGTATTCGCGGTA 58663
QY 272 TCCAGCCAGTTACAGCGTGTACTGATTTGATGTTGTTGTCAGCGGGAAGCAGTACG 331
DB 58664 TGGCGGCTCAATATACAAATGATTTGGTTGATGTTAAGCGTGTGATACCGGAGTACTC 58723
QY 332 TGACTCCCAACGGTTTTTCTGCCATGAATACCGGTTTCATGCCCTCTTGGCCGCCATTG 391
DB 58724 GCCCAACCGTGATAATTCGGGGATTGAACAGGGGTGTTGCCGCGCATTCGCCGATAG 58783
QY 392 AGCGTATTAGGTTATCAGGGGGCGGATGTCCACACTGTATGGCTCTGATCCGATGGCGG 451
DB 58784 AACGTATCAGGTTGTTCTGTGCTCTATGCTCTTCTTTATGTTCTGTATGCGATGGCG 58843
QY 452 GTGTGGTGAATATCATTTACCAAGAAAGATGCAGACAAATGGCTCTCTTCCTCAATCAG 511
DB 58844 GGGTGAATTAACATCATTTACCGGTAAAGCAGAGAAGATGGAACACAGCTTCGCCGCTG 58903
QY 512 GGCTGAATCTGCAGAAAAGCAAAATGGGGTAAACAGCAGCCAGTTTAATTTCTGGAGCA 571
DB 58904 ATGGCACTTTGCAGGAGGTTCTTAATCAGGTAATAGTCAATCAGAGTAGTGTGTATACTT 58963
QY 572 GTGTGTCCTCTGTGATGATTTCTGTCAGCTGAGGTACGCGG 614
DB 58964 CAGGTCCAATATTATTGATGATTTAGGGCTGAAAGTGAGCGG 59006

RESULT 7

ACF67367_01
Continuation (2 of 57) of ACF67367 from base 100001 (Photorhabdus luminescens nucleotide
WP Sequence split into 57 fragments LOCUS ACF67367 Accession ACF67367
WP Fragment Name Begin End
ACF67367_01 1 110000
ACF67367_02 100001 210000
ACF67367_03 200001 310000
ACF67367_04 300001 410000
ACF67367_05 400001 510000
ACF67367_06 500001 610000
ACF67367_07 600001 710000
ACF67367_08 700001 810000
ACF67367_09 800001 910000
ACF67367_10 900001 1010000
ACF67367_11 1000001 1110000
ACF67367_12 1100001 1210000
ACF67367_13 1200001 1310000
ACF67367_14 1300001 1410000
ACF67367_15 1400001 1510000
ACF67367_16 1500001 1610000
ACF67367_17 1600001 1710000
ACF67367_18 1700001 1810000
ACF67367_19 1800001 1910000
ACF67367_20 1900001 2010000
ACF67367_21 2000001 2110000
ACF67367_22 2100001 2210000
ACF67367_23 2200001 2310000
ACF67367_24 2300001 2410000
ACF67367_25 2400001 2510000
ACF67367_26 2500001 2610000
ACF67367_27 2600001 2710000
ACF67367_28 2700001 2810000
ACF67367_29 2800001 2910000

WP	ACF67367_29	2900001	3010000
WP	ACF67367_30	3000001	3110000
WP	ACF67367_31	3000001	3210000
WP	ACF67367_32	3200001	3310000
WP	ACF67367_33	3300001	3410000
WP	ACF67367_34	3400001	3510000
WP	ACF67367_35	3500001	3610000
WP	ACF67367_36	3600001	3710000
WP	ACF67367_37	3700001	3810000
WP	ACF67367_38	3800001	3910000
WP	ACF67367_39	3900001	4010000
WP	ACF67367_40	4000001	4110000
WP	ACF67367_41	4100001	4210000
WP	ACF67367_42	4200001	4310000
WP	ACF67367_43	4300001	4410000
WP	ACF67367_44	4400001	4510000
WP	ACF67367_45	4500001	4610000
WP	ACF67367_46	4600001	4710000
WP	ACF67367_47	4700001	4810000
WP	ACF67367_48	4800001	4910000
WP	ACF67367_49	4900001	5010000
WP	ACF67367_50	5000001	5110000
WP	ACF67367_51	5100001	5210000
WP	ACF67367_52	5200001	5310000
WP	ACF67367_53	5300001	5410000
WP	ACF67367_54	5400001	5510000
WP	ACF67367_55	5500001	5610000
WP	ACF67367_56	5600001	5648894

XX	AAS46270;		
XX	18-DEC-2001 (first entry)		
DT	DNA encoding novel mar regulated protein (NIMR) #39.		
DE	mar regulated polypeptide; NIMR; microbial infection; antibacterial; ds.		
XX	Escherichia coli.		
KW	WO200170776-A2.		
XX	27-SEP-2001.		
XX	08-MAR-2001; 2001WO-US0007478.		
PF	10-MAR-2000; 2000US-0188362P.		
XX	(TUFT) TUFTS COLLEGE.		
XX	Levy SB, Barbosa TM, Alekshun MN;		
XX	WPI; 2001-602769/68.		
DR	P-PSDB; AAU29371.		
XX	Identifying compounds that modulate a newly identified mar regulated		
PT	polypeptide activity, useful as antimicrobial compounds, involves		
PT	contacting the polypeptide with a test compound.		
XX	Disclosure; Page 451-457; 526pp; English.		
PS	The invention relates to a method of identifying compounds that modulate		
CC	a newly identified mar regulated (NIMR) polypeptide activity. The method		
CC	comprises contacting an NIMR polypeptide with a test compound under		
CC	interaction conditions, determining the ability of the compound to		
CC	modulate the activity or expression of the polypeptide, and selecting the		
CC	modulators. NIMR nucleic acids and polypeptides are used in the treatment		
CC	of microbial infections, and in screening for modulators of NIMR		
CC	expression and activity. These modulators can be used to reduce the		
CC	infectivity of a microbe on a surface, and the virulence of a microbe in		
CC	a subject suffering from an infection. AAS46232-AAS46278 represent		
CC	Escherichia coli NIMR coding sequences of the invention		
XX	Sequence 10353 BP; 2559 A; 2763 C; 2485 G; 2546 T; 0 U; 0 Other;		
SQ	Query Match 6.2%; Score 130.6; DB 4; Length 10353;		
	Best Local Similarity 54.6%; Pred No. 3.4e-30;		
	Matches 312; Conservative 0; Mismatches 244; Indels 15; Gaps 2;		
QY	62 TAGCTGCTCAGAGGATGTGATGGTTCGGCATCCGGTATGAGAAAAGTGACTA 121		
Db	10102 TCGATGATGTCGGAACGATGGTTGTCACTGCATCTTCGGTGAAACAAATCTTAAG 10043		
QY	122 ACCACGCCCGAGTGTTCGTGATTAGCCAGRGGAATTCAGTCCAGCCAGTACCACG 181		
Db	10042 ATGCACCTCGCAGTATCACCGTCATTCACAGGAAGACCTCGAGGAAACCCGTACAGA 9983		
QY	182 ATCTGCGGAGGCTCTGAGATCAGTAGAGGGTGTGGATGTTGAAAGTGGTACGGGTAAAA 241		
Db	9982 ATCTGAAGGATGTCTCTCAAAGAAAGTGCCTGGCGTACAACACTGACGAACGAGGGGATAA-- 9925		
QY	242 CCGAGGGCTGGAATCAGCATCCGAGGAATCCAGGAGTACACCGTGTATCTGATTG 301		
Db	9924 ----CCGTAAAGGCGGTAGTATTCGTGTGTGGACAGCAGCTATACCCCTGATTCTCTGTCG 9869		
QY	302 ATGGTGTTCGTACGGCGGGAAGCAGTGCAGTGCATCTCCCAACGGTTTTTCTGCCATGAATA 361		
Db	9868 ACGGTAAAACGCGTGAACCTCCGCAATCCGCTCTCCGCCAACATGATTCATCTGAA-- 9811		
QY	362 CGCGGTTCAATGCCCCCTCTGCGCGCGCATTTAGCGTATTTAGCGTTATCAGGGGGCGGATGT 421		
Db	9810 -----CTGGATCCCGGTCCGATTCATCGAACGTATTGAAGTGGTCCGTGCGCCCGATGT 9758		

WP	ACF67367_29	2900001	3010000
WP	ACF67367_30	3000001	3110000
WP	ACF67367_31	3000001	3210000
WP	ACF67367_32	3200001	3310000
WP	ACF67367_33	3300001	3410000
WP	ACF67367_34	3400001	3510000
WP	ACF67367_35	3500001	3610000
WP	ACF67367_36	3600001	3710000
WP	ACF67367_37	3700001	3810000
WP	ACF67367_38	3800001	3910000
WP	ACF67367_39	3900001	4010000
WP	ACF67367_40	4000001	4110000
WP	ACF67367_41	4100001	4210000
WP	ACF67367_42	4200001	4310000
WP	ACF67367_43	43	

RESULT 8

AAS46270/c

ID AAS46270 standard; DNA; 10353 BP.

422 CCACACTGTATGGCTCTGATCGATGGCGGTGGTGGATATCATTTACCAGAAAGATG 481
 9757 CGTCGCTGTACGGTTCGATGGCGGTGGTGGATATCATTTACCAGAAAGATG 9698
 482 CAGACAAATGGCTCTCTCCGTCGAATGACAGGGTGAATTCGAGGAAAGCAACAATGGG 541
 9697 GTACAGAAATGGCTGGGTACCGTTACCGTCGATACCAACCAATTCAGGAACATCGCGATCGG 9638
 542 GTACACACAGCCAGTTTAAATTTCTGGAGCAGTGGTCCCTTGGGATGATCTGTCAGCC 601
 9637 GTGACACCTTAACCGTCACTTTTACCAAGTGAACCAATTAATGATGGTGTCTGGAA 9578
 602 TGCAGGTACGCGGTAGCACACAACAGCGTCA 632
 9577 TGAAGCTTACGCGCGCTGGCAAAAGCTGA 9547
 RESULT 9
 ABA89269
 ID ABA89269 standard; DNA; 2046 BP.
 AC ABA89269;
 DT 11-FEB-2002 (first entry)
 DE Escherichia coli polynucleotide SEQ ID NO 1076.
 KW Escherichia coli; B2/D-A; antiinflammatory; antibacterial;
 KW immunosuppressive; extra-intestinal infection; phylogeny; meningitis;
 KW systemic infection; non-diarrhoeal infection; septicemia;
 KW pyelonephritis; antibiotic resistance; ds.
 OS Escherichia coli.
 WO200166572-A2.
 13-SEP-2001.
 12-MAR-2001; 2001WO-BP003445.
 10-MAR-2000; 2000FR-00003145.
 02-FEB-2001; 2001FR-00001449.
 (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 Bingen E, Bonacorsi S, Clermont O, Nassif X, Tinsley C;
 WPI; 2001-550253/61.
 A library of DNA fragments of Escherichia coli strains for the phylogenetic determination of a given strain comprises polynucleotides of nature B2/D+ A-.
 Example 6; Fig 6; 646pp; English.
 The invention relates to a library of DNA fragments of Escherichia coli strains comprising polynucleotides (ABA88577-ABA88729 and ABA89533) and encoded proteins (AB52459-AB52919 and AB52954-AB53094) of nature B2/D+ A- . The polynucleotides have potential antiinflammatory, antibacterial and immunosuppressive activity as part of pharmaceutical compositions used to treat, palliate or prevent extra-intestinal E. coli infections. The polypeptides are useful for determining the phylogenetic group of a given E. coli strain. These polypeptides can detect and treat an undesired development of E. coli, particularly an extra-intestinal infection that include systemic and non-diarrhoeal infections such as septicemia, pyelonephritis and meningitis this is particularly advantageous as bacterial resistance is increasing with the more frequent use of broad spectrum antibiotics
 Sequence 2046 BP; 645 A; 396 C; 484 G; 521 T; 0 U; 0 Other;
 Query Match 6.0%; Score 125.6; DB 4; Length 2046;

Best Local Similarity 46.3%; Pred. No. 5.4e-29;
 Matches 684; Conservative 0; Mismatches 759; Indels 33; Gaps 7;
 QY 72 AGAGATGTGATGATGTTGCTCGGCATCGGCTATGAGAAAAGCTGACTAACGAGCGCG 131
 DB 87 AGAAGATAGCTGTTGTTGTTACTGCTCGGTTCACTCAGCAGCTCAGAATGCCCGGC 146
 QY 132 CAGTGTCTCTGTGATTAACCCAGGAGGAATGCACTCCAGCAGTACCAAGATCTGGCGGA 191
 DB 147 CAGTGTCTCACTCATCTTCCAGAACCACTGCAAAAAAACCCTTTTCAGATCTGGTCTGA 206
 QY 192 GGCTCTGAGATCAGTAGAGGGTGGATGTTGAAGTGGTACGGGTAAACCCGAGGGCT 251
 DB 207 TG-----CAGTAAAGATGTTGAAGGATTAGTATCACTGGTGGGAATGAAAAACC 257
 QY 252 GGAATCAGCATCCGAGGAATGCCAGCCAGTACACGCTGATGATGATGATGTTGTTG 311
 DB 258 GGAATCAGTATACGTGCTTAAGTGGCATTTACACGCTGATTCGTCGATGGACGACG 317
 QY 312 TCAGGCGGAAGCAGTAGTACGCTCCCAACGGTTTTCTGCCATGATATACCGGTTTCA 371
 DB 318 TCAGAGCG---GTCCGGGAATCCAGCAAAACCGGCGGGTTTTTGAAGCCGGAATTTAT 374
 QY 372 GCCCCTCTGGCGGCATTGAGCGTATTGAGTTTATCAGGGGGCGGATGCCACACTGTA 431
 DB 375 CCTCTCTGGAAGCAATTGAACGATTTGAAGTATCGGTGGCCCTATGTTCTCCCTGTA 434
 QY 432 TGGCTCTGATCGGATGGCGGTGGTGAATATATATACAGAAAGAAATGCAGACAAATG 491
 DB 435 TGGTTCTGATGCCATCGAGGGGTCAATTAATATATACCAAAACCACTTAATAACCAAA 494
 QY 492 GCTCTCTCCGTCATGCA---GGGCTGAATCTCAGGAAAGCAACAAATGGGGTAAACAG 548
 DB 495 ATGGGATGGCGTACTTGGACTTTGGGGGATTTTCAGGAACATGGGAAATTTGGTAATC 554
 QY 549 CAGCCAGTTTAAATTTCTGGAGCAGTGTCCCTTGTGGATGATTTCTGACGCTGACGTT 608
 DB 555 AACCAAAATGACTTCTATCTGTCAGGCCCATTTGATTAAGGATAAACTTGGTCTTCACT 614
 QY 609 ACGGGTAGCACACAACAGCTCAGGGTTTCATCGTCACTCACTGAGCGCATACAGCAGG 668
 DB 615 ATATGGAGGAATGAACATATCCAAAGG-----AGATAGTATCTCTCAGGA-AACCGCG 668
 QY 669 CACGCTATTCTCTTATCCACGGAGTCAAGAAATTAATCTTGGTGCACGCTCTTGAATG 728
 DB 669 AAAGATAAATAAGAAATATAACGGCAACGCTCCAGTTTACTCCGACTGAAAGCCAGAAGTT 728
 QY 729 GAAGCGTCCGAGCAGGATGTGCTCGTGTGATATGATACCAACCCGAGCGGTTATGA 788
 DB 729 TGTTTTGAATATGAAAAAATAACAGGTGCATACATTAACACCTGGTGAATCTCTCGA 788
 QY 789 TAAACGGGATGGCAACTGGGGAGTCTGACGGGGGATATGACCGGACCTTCGCTATGA 848
 DB 789 TGCCTGACTATGCGGGGAATCTTAACACCAACAGTAAGAAAGAAACGCAATATTC 848
 QY 849 CGCAACAAATTTTCACTGGCTATGATCATCTTTCACCTTCGGAACATGGAATCGTA 908
 DB 849 ACCTAGTCACTCGGTAGCAGCAT-----GGAATGCCCGCGGCAAAATCTGCACTCTGA 902
 QY 909 TCTGAACCTGGAACGAGACAGAAAAATAAAGTTCGTGAGCTTGTACGCACTGTACTGAAGCG 968
 DB 903 AATTGCTGTTTATCAGGAGAAAGTTTATTCGTGAGTTAATCAGTAAAAAGATAATA 962
 QY 969 CGCAAAATGGGGCTTCCCGTACGCGCGGAGCTTAAGGAATCGAAACCTTATCTGAA 1028
 DB 963 TAATCATTTGGGATCTTAATTAACGAGTCAAGAAAAACCGAAAAATAACCAACAATCATAGA 1022
 QY 1029 TCAATTAAGTACCTTCCCTCTGGGAGATCTCATCTGTTACGGTGGGGCGGATTTCA 1088
 DB 1023 TGAAGAGTGAAGGATTTTCTGCCGGAATTTGACTGACCATCGGAGTCAATTTACGA 1082
 QY 1089 GAGCTCGTCCATGAAGACGAGTTGCTTTCGCCAGCACAGGTGAAA-----CTTTCCG 1142

Db 1083 TGCAGAGCTCCGTGATGACTCAGCCCGGTTAAAAAAGCAGACAGAAACACAGTCTGTTC 1142
 Qy 1143 GCAGAAAAGCTGGTGGTATTTCTGAGGATGAGTGGCATCTCAGGATGCACTTGGCGT 1202
 Db 1143 AATTAACACAGAGCTGTTTATAGAAATGAATATGACGAACGGATTCTCGCCCT 1202
 Qy 1203 GACTGGGGCAGCGGCTATGAACATCATGAGCAATTCGGGGGACACTTCAGTCCGGTGC 1262
 Db 1203 GACTGGAGGAGCTCGTCTCGATAATCATGAAATCTATGGCAGTTACTGGAATCCAAGATT 1262
 Qy 1263 ATATCTGCTGGATGTGACAGATGCTGGACGCTGAAAGGGGGTGTGACACGGGATA 1322
 Db 1263 GTAGCTGTTTAACTGACCGATTAATCTCACATCAAGAGGGGATGCAAAAGCATT 1322
 Qy 1323 TAAGGACCCAGAAATGGGCGAGTACATAAAGGATTAAGTGTGTCGCGGACAGGAAA 1382
 Db 1323 TCGGGCTCTTCAATTCGTGAGTGAGTCTGGATTGGAACACTGACGACGGTGGTGC 1382
 Qy 1383 AACAAATCTACTTGTAAACCCGACCTGAGCGGAGAGAGCGTCAAGTTATGAGCTGG 1442
 Db 1383 CTCTATTATGATGGAACAGAGGACCTGAAACCGGAGACCGATGTAACCGAAGATCGG 1442
 Qy 1443 GGTGTATTACGATAAACCCCGCGCTCTGAATGCAATGTCAACAGGTTTATGACTGACTT 1502
 Db 1443 TATTATTATGATGATAGTGTGTTTTCGGGAGCGGACGCTGTTTAACTGATT 1502
 Qy 1503 CTCACACAGATGTCTCTTATTCATTAATGATAA 1538
 Db 1503 TAAAAATAAGTTGACCAGTTACGATATAGTACAAA 1538

RESULT 10

AB89255 standard; DNA; 12943 BP.
 AC AB89255;
 XX
 DT 11-FEB-2002 (first entry)
 XX
 DE Escherichia coli polynucleotide SEQ ID NO 1049.
 XX
 KW Escherichia coli; B2/D+A-; antiinflammatory; antibacterial;
 KW immunosuppressive; extra-intestinal infection; phylogeny; meningitis;
 KW systemic infection; non-diarrhoeal infection; septicaemia;
 KW pyelonephritis; antibiotic resistance; ds.
 XX
 OS Escherichia coli.
 XX
 PN WO200166572-A2.
 XX
 PD 13-SEP-2001.
 XX
 PF 12-MAR-2001; 2001WO-EF003445.
 XX
 PR 10-MAR-2000; 2000FR-00003145.
 PR 02-FEB-2001; 2001FR-00001449.
 XX
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 XX
 PI Bingen E, Bonacorsi S, Clermont O, Nassif X, Tinsley C;
 XX
 DR WPI; 2001-550253/61.
 XX
 PT A library of DNA fragments of Escherichia coli strains for the phylogenetic
 PT determination of a given strain comprises polynucleotides of nature B2/D+
 PT A-.
 XX
 PS Example 6; Fig 6; 646pp; English.
 XX
 CC The invention relates to a library of DNA fragments of Escherichia coli
 CC strains comprising polynucleotides (AB89257-AB89279 and AB89533) and
 CC encoded proteins (AB89245-AB89291 and AB89294-AB89304) of nature
 CC B2/D+A-. The polynucleotides have potential antiinflammatory,
 CC B2/D+A-.

CC antibacterial and immunosuppressive activity as part of pharmaceutical
 CC compositions used to treat, palliate or prevent extra-intestinal E. coli
 CC infections. The polypeptides are useful for determining the phylogenetic
 CC group of a given E. coli strain. These polypeptides can detect and treat
 CC an undesired development of E. coli, particularly an extra-intestinal
 CC infection that include systemic and non-diarrhoeal infections such as
 CC septicemia, pyelonephritis and meningitis this is particularly
 CC advantageous as bacterial resistance is increasing with the more frequent
 CC use of broad spectrum antibiotics
 XX

SQ Sequence 12943 BP; 3358 A; 2856 C; 3166 G; 3563 T; 0 U; 0 Other;

Query Match 6.0%; Score 125.6; DB 4; Length 12943;
 Best Local Similarity 46.3%; Pred. No. 1.6e-28;
 Matches 684; Conservative 0; Mismatches 75; Indels 33; Gaps 7;
 Qy 72 AGAGATGTGATGTTGTCTCGCATCGGCTATGAGAAAAGCTGACTAACGAGCGC 131
 Db 10799 AGAAGATAGCTTGTGTTACTGCTCGGTTCACTCAGCAGCTCAGAATGCCCGC 10858
 Qy 132 CAGTGTCTGTGATTAGCCAGAGGAATTCAGTCCAGCCAGTACCAGATCTGGCGGA 191
 Db 10859 CAGTGTCTCAGTCATTACTTCAGAACAACTGCAAAAAAACCCTTCAGATCTGGTGA 10918
 Qy 192 GCCTCTGAGTCACTAGAGGCTGTGGATGTTGAAAGTGGTACGGTAAACCGGAGGCT 251
 Db 10919 TG-----CAGTAAAGATGTTGAAGGATTAGTATCACTGGTGGATGAAAC 10969
 Qy 252 GGAATCAGCATCCGAGGAATGCCAGCAGTTACGCTGATGATGATGATGTTGTCG 311
 Db 10970 GGATATCAGTATAGTGTCTAAGTGGCGATTACGCTGATCTCTGGTCGATGACGACG 11029
 Qy 312 TCAGGGCGAGCAGTGAAGTGTCTCCCAAGGTTTTCTGCCATGATACCGGTTTCA 371
 Db 11030 TCAGAGCG---GTCGGGAATCCAGACCAACCGGAGCGCGGTTTTGAAGCGGATTTAT 11086
 Qy 372 GCGCCCTCTGCGCCCATTTGAGCGTATTGAGTTTACAGGGGCGCGATGTCCACATGTA 431
 Db 11087 CCTCCTGTGGAAGCAATTGAACGATTGAAGTATCGTGGCCCTATGTCTCCCTGTA 11146
 Qy 432 TGGCTCTGATCGGATGGCGGTGTGGTGAATATCATTTACAGAAAGATGAGACAAATG 491
 Db 11147 TGGTCTGATCCATCGGAGGGGTCAATATATATCAACCAACAGTAAATACCAAC 11206
 Qy 492 GCTCTCTCCGTCAATGCA---GGGCTGAATCTGAGGAAAGCAACAAATGGGTAAACAG 548
 Db 11207 ATGGGATGGCTACTTGGACTTGGGGGATTTTCAGGAACATGGGAAATTTGGTAATC 11266
 Qy 549 CAGCCAGTTTAAATTTCTGGAGCAGTGTCCCTTGTGGATGATTTCTGTCAACCTGCGAGG 608
 Db 11267 AACCAAAATGACTTCTATCTGTGAGGCCCATTTGATTAAGGATAAACTTGTCTTCAGCT 11326
 Qy 609 ACGCGTAGCACACAGAGCGTCAAGGTTTCATCGTCACTCAGCGGATACAGCAGG 668
 Db 11327 ATATGGAGATGAATCTATCCAAAG-----AAGATAGTATCTCTCAGGGA-ACACGGC 11380
 Qy 669 CACCGTATCTCTTATCCACGGAGTCAAGAAATTAATCTTGGTGCACGCTTTGACTG 728
 Db 11381 AAAAGATAATAAGAAATATAACGCAACGCTCCAGTTTACTCCGACTGAAAGCCAGAACTT 11440
 Qy 729 GAAGCGTCGAGCAGGATGTCTCTGTTTGTATATGATACACCGGAGCGGAGCGCTTATGA 788
 Db 11441 TGTTTTGAAATATGAAAAAATAAACAGGTGCATACATTAAACCTGGTGTCTCTCGA 11500
 Qy 789 TAACCGGATGGCAACTGGGAGTCTGAACGGGGGATATGACCGGAGCCCTTGGCCTATGA 848
 Db 11501 TGCCTGGACTATCGGGGAATCTTAAACACCAACAGTAAGAGAAAGCATAATTC 11560
 Qy 849 GCGAAACAAATTCAGCTGGCTATGATCATCTTTCACCTTCGGAAACATGGAATCGTA 908
 Db 11561 ACGTAGTCACTGGGTAGCAGCAT-----GGAATGCCAGGGCGGAAATCTGTCATCTGA 11614
 Qy 909 TCTGAAGTGGACGAGACAGAAATAAAGTGTGAGCTTTGTACGAGTGTACTGAAGCG 968

CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS84157-AAS94564 represent novel human diagnostic
 CC coding sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 3675 BP; 914 A; 1051 C; 975 G; 735 T; 0 U; 0 Other;
 Query Match 3.3%; Score 69.4; DB 5; Length 3675;
 Best Local Similarity 57.7%; Pred. No. 1.1e-10;
 Matches 124; Conservative 0; Mismatches 91; Indels 0; Gaps 0;
 QY 418 ATGTCACACATGATGCTCTGATGCGATGCGGGTGTGGTGAATATCATACCAAG 477
 DB 1 ATGTCGTCGTGACGGTTCGATGCGTGGGGTGTAGTGAATATCATACCAAG 60
 QY 478 AATGACAGCAAAATGGCTCTCTTCGTCATGTCAGGGCTGAATCTGAGGAAGCAAA 537
 DB 61 ATCGGTGAGAAATGGTGGGTACCGTTACCGTCGATACCAATCAGGAATCCGAT 120
 QY 538 TGGGGTACACGACCGAGTTAATCTGAGAGAGTGGTCCCTGTGGATGATTCGTC 597
 DB 121 CGCGGTACACCTATACGGTACGTTCTTTACAGTGGACCAATTAATGATGGTGTGCTG 180
 QY 598 AGCCTGACGGTACGGGTAGCACACACAGCGTCA 632
 DB 181 GGAATGAAGCTTACGGCAGCCTGGCAACAGTGA 215

RESULT 14
 AAD03054
 ID AAD03054 standard; DNA; 2495 BP.
 AC AAD03054;
 DT 13-JUN-2001 (first entry)
 XX ironNec gene fragment from an extraintestinal isolate of Escherichia coli.
 XX ironNec; extracytoplasmic protein; immunogen; vaccine; BI; UTI;
 XX immunotherapy; extraintestinal infection; urinary tract infection;
 XX meningitis; pneumonia; intra-abdominal infection; antibiotic; ds.
 XX Escherichia coli.
 OS
 XX Key Location/Qualifiers
 XX CDS 321..2495
 XX /tag= a
 XX /product= "E. coli ironNec extracytoplasmic protein"
 XX /transl_except= (pos:678..686, aa:Val-Tyr)
 XX /note= "Insertion of 3 bases alters the reading frame"
 XX /transl_except= (pos:723..728, aa:Asn-Arg-Tip)
 XX /note= "There is an apparent deletion of one codon
 XX corresponding to Arg, which alters the reading frame; the
 XX coding region does not include stop codon"
 XX /partial
 XX 321..392
 XX sig_peptide
 XX /tag= b
 XX 393..2495
 XX mat_peptide
 XX /tag= c
 XX /product= "Mature E. coli ironNec extracytoplasmic

FT protein"
 XX WO2000121636-A1.
 PN 29-MAR-2001.
 PD 22-SEP-2000; 2000WO-US026117.
 XX 22-SEP-1999; 99US-0155621P.
 PR (UUNY) UNIV NEW YORK STATE RES FOUND.
 PA Russo T, Carlino U;
 PI WPI; 2001-244936/25.
 DR P-PSDB; AAY72914.
 XX Novel isolated ironNec polynucleotide from extraintestinal isolate of
 PT Escherichia coli useful as vaccine for treating or preventing
 PT extraintestinal infections caused by extraintestinal pathogenic
 PT Escherichia coli.
 XX Claim 5; Page 37-38; 44pp; English.
 CC The invention relates to the identification of ironNec gene, from an
 CC extraintestinal isolate of Escherichia coli. This gene is expressed in
 CC increased amounts in human urine and is identified by transposon (TnpA)
 CC mutagenesis. ironNec gene encodes an extracytoplasmic protein. This gene
 CC can be used as an immunogen in vaccine formulations. The recombinant
 CC vector comprising nucleotide sequence encoding one or more antigenic
 CC epitope of ironNec is useful for diagnostic and immunotherapeutic
 CC purposes. The ironNec antigenic peptide is useful for treating or
 CC preventing extraintestinal infections (BIs) caused by extra- intestinal
 CC pathogenic E. coli (ExPEC). The BIs include urinary tract infection
 CC (UTI), meningitis, intra-abdominal infection and pneumonia. The present
 CC sequence is ironNec gene, from an extraintestinal isolate of Escherichia
 CC coli
 XX
 SQ Sequence 2495 BP; 686 A; 557 C; 632 G; 620 T; 0 U; 0 Other;
 Query Match 3.3%; Score 68.6; DB 4; Length 2495;
 Best Local Similarity 55.1%; Pred. No. 1.1e-10;
 Matches 134; Conservative 0; Mismatches 109; Indels 0; Gaps 0;
 QY 376 CCTCTGCGCCCATGAGCGTATGAGGTATTCAGGGGGCGGATGTCACACTGTATGCG 435
 DB 732 CCACGGCAACAGTTGAGCGTATTCAGCGGCGCTCGCGGGCGCGGTACGGT 791
 QY 436 TCTGATGCGATGGCGGTGTGTCATATCATATTACCAAGAAAGATGCAGACAAATGGCTC 495
 DB 792 TCGGGGGCGCGGGGGGTGTGTAACATCATATTACCAACGTCGCCACACAGCTGGCAC 851
 QY 496 TCTTCCCGTCAATGACAGGCTGAATCTGCAGAAAGCAACAAATGGGTACAGCAGCGAG 555
 DB 852 GGTTCGCTGTGCTTATACACCAACAGCCGCGAAAGTAGCGAAGAGGGCGGTACGCGTCG 911
 QY 556 TTTAATTTCTGAGCAGTGGTCCCTTGTGGATCATCTGTGACGCTCAGGTACGGGT 615
 DB 912 GCCAATTTTCAGCCCTTAGTGGGCGCTCTGGCTGGTATGCTCTTACACGCGTTGTATGGT 971
 QY 616 AGC 618
 DB 972 AAC 974
 XX
 XX RESULT 15
 XX AAL15184/c
 XX ID AAL15184 standard; DNA; 4039 BP.
 XX AAL15184;
 XX AC AAL15184;
 XX DT 04-SEP-2000 (first entry)
 XX

DE DNA encoding Escherichia coli virulence proteins.

KW Virulence protein; tatA; tatB; tatC; tatE; mdoG; creC; recG; yggN; eckI;
KW iroD; iroC; iroE; mtd2; msl; vaccine; infection; Gram negative Bacterium;
KW ss.

XX Escherichia coli.

XX Key Location/Qualifiers

XX CDS 1..285

XX FT /*tag= a

XX FT /product= "virulence protein"

XX FT /note= "encodes AAY93235"

XX CDS 370..1326

XX FT /*tag= b

XX FT /product= "virulence protein"

XX FT /note= "encodes AAY93236"

XX PN WO200028038-A2.

XX PD 18-MAY-2000.

XX PF 09-NOV-1999; 99WO-GB003721.

XX PR 09-NOV-1998; 98GB-00024569.

XX PR 09-NOV-1998; 98GB-00024570.

XX PR 17-DEC-1998; 98GB-00027814.

XX PR 17-DEC-1998; 98GB-00027815.

XX PR 17-DEC-1998; 98GB-00027816.

XX PR 17-DEC-1998; 98GB-00027818.

XX PR 13-JAN-1999; 99GB-00000708.

XX PR 13-JAN-1999; 99GB-00000710.

XX PR 13-JAN-1999; 99GB-00000711.

XX PR 28-JAN-1999; 99GB-00001915.

XX PA (MICR-) MICROSCIENCE LTD.

XX PI Crooke HR, Clarke EE, Everest PH, Dougan G, Holden DW, Shea JE;

XX PI Feldman RG;

XX DR WPI; 2000-376550/32.

XX DR P-PSDB; AAY93235, AAY93236.

XX PT Peptide encoded by an operon including genes from Escherichia coli for

XX PT screening potential drugs, detecting virulence and treating conditions

XX PT associated with infection by a Gram negative bacterium.

XX PS Example 10; Page 74-78; 122pp; English.

XX CC The present sequence encodes Escherichia coli virulence proteins. The

XX CC specification describes virulence proteins which are encoded by an operon

XX CC including tatA, tatB, tatC, tatE, mdoG, creC, recG, yggN, eckI, iroD,

XX CC iroC, iroE, mtd2 or msl-15 genes obtained from Escherichia coli K1. The

XX CC virulence proteins and polynucleotides, and their vaccines are useful for

XX CC screening potential drugs, for the detection of virulence, and for

XX CC treating or preventing conditions associated with infection by a Gram

XX CC negative bacterium particularly Escherichia coli

XX SQ Sequence 4039 BP; 985 A; 1039 C; 942 G; 1073 T; 0 U; 0 Other;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 12, 2004, 11:56:38 ; Search time 8116 Seconds
(without alignments)
11166.866 Million cell updates/sec

Title: US-10-625-972-4
Perfect score: 2091
Sequence: 1 atcgagaataacactctggc.....cgctgaactatcagttcttga 2091

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.ov.*
- 22: em.pat.*
- 23: em.ph.*
- 24: em.pl.*
- 25: em.ro.*
- 26: em.sts.*
- 27: em.un.*
- 28: em.vi.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.lrv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.rod.*
- 36: em.htg.nam.*
- 37: em.htg.vit.*
- 38: em.sy.*
- 39: em.hgo.hum.*
- 40: em.hgo.mus.*
- 41: em.hgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	2091	100.0	2091	6	AR025458	Sequence
2	2091	100.0	8041	6	AR025455	Sequence
3	2089.4	99.9	8040	1	AF126104	Escherich
C 4	2089.4	99.9	10029	1	AE005310	Escherich
C 5	2089.4	99.9	10155	1	AE005274	Escherich
C 6	2089.4	99.9	87583	6	AR204161	Sequence
C 7	2087.8	99.8	86248	6	BD184766	Nucleic a
C 8	2087.8	99.8	327773	1	AP002554	Escherich
C 9	2079.8	99.5	76804	6	AX702523	Sequence
C 10	2079.8	99.5	303121	1	AF016766	Escherich
C 11	2046.8	97.9	13710	1	AF081285	Escherich
C 12	1863.8	89.1	37710	1	ECO278144	Escherich
C 13	1025	49.0	71684	1	AF447814	Escherich
C 14	646.4	30.9	10029	1	AE013687	Yersinia
C 15	644.8	30.8	220050	1	AJ414156	Yersinia
C 16	177	8.5	343473	1	EX640451	Bordetella
C 17	175.4	8.4	3169	1	BBU56084	Yersinia
C 18	174.4	8.3	300507	1	AE017146	Helicobac
C 19	173.8	8.3	1239	1	AY212670	Unculture
C 20	153.8	7.4	317511	1	CJ11168X3	Campyloba
C 21	150	7.2	3300	1	CCU80812	Campylobact
C 22	149.4	7.1	345829	1	EX571869	Photornab
C 23	149.4	7.1	349980	6	AX770905	Sequence
C 24	147.8	7.1	10354	1	AE013889	Yersinia
C 25	147.8	7.1	199050	1	AJ414147	Yersinia
C 26	142.6	6.8	312839	1	AE016858	Pseudomon
C 27	137.8	6.6	12635	1	AE009322	Agrobacte
C 28	137.8	6.6	13095	1	AE008292	Agrobacte
C 29	134.2	6.4	3339	1	AE052655	Vibrio pa
C 30	134.2	6.4	296650	1	AP005082	Vibrio pa
C 31	130.6	6.2	2499	1	ECOCIR	E.coli coli
C 32	130.6	6.2	10342	1	AE005447	Escherich
C 33	130.6	6.2	10353	1	AE000304	Escherich
C 34	130.6	6.2	10353	6	AX370264	Sequence
C 35	130.6	6.2	75888	1	ECOH047	Sequence
C 36	130.6	6.2	268857	1	AP002560	Escherich
C 37	129	6.2	12697	1	AE015239	Shigella
C 38	129	6.2	295689	1	AE016985	Shigella
C 39	127	6.1	2055	6	AR375730	Sequence
C 40	125.6	6.0	2046	6	AX276467	Sequence
C 41	125.6	6.0	2046	6	AX824854	Sequence
C 42	125.6	6.0	2046	6	AX828804	Sequence
C 43	125.6	6.0	2049	1	AF320691	Escherich
C 44	125.6	6.0	12943	6	AX276440	Sequence
C 45	125.6	6.0	306118	1	AE016771	Escherich

ALIGNMENTS

RESULT 1	AR025458	2091 bp	DNA	linear	PAT 05-DEC-1998
LOCUS	Sequence 4 from patent US 5798260.				
DEFINITION	Sequence 4 from patent US 5798260.				
ACCESSION	AR025458				
VERSION	AR025458.1				
KEYWORDS	GI:3978086				
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	Unclassified.				
AUTHORS	1 (bases 1 to 2091)				
TITLE	Tarr, P.I., Bilge, S.S., Besser, T.E. and Vary, J.C. Jr.				
JOURNAL	Escherichia coli O157:H7 epithelial adhesin				
FEATURES	Patent: US 5798260-A 4 25-AUG-1998;				
	Location/Qualifiers				

source	1. .2091																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
--------	----------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

QY	961	CTGAAGCGCGACAAATGGGGCTTTCGGGTACGCGCGGGAGCTTAAGGAATCGAACCTT	1020			
DB	961	CTGAAGCGCGACAAATGGGGCTTTCGGGTACGCGCGGGAGCTTAAGGAATCGAACCTT	1020			
QY	1021	ATCCTGAATTCATTACTGTCTTACCCCTCTGGGAGAAATCTCATCTGTTACGGTGGGGGC	1080			
DB	1021	ATCCTGAATTCATTACTGTCTTACCCCTCTGGGAGAAATCTCATCTGTTACGGTGGGGGC	1080			
QY	1081	GAGTTTCAGAGCTCGTCCATGAAAGACGAGTGTCTCTGCCAGCACAGGTGAAACTTTTC	1140			
DB	1081	GAGTTTCAGAGCTCGTCCATGAAAGACGAGTGTCTCTGCCAGCACAGGTGAAACTTTTC	1140			
QY	1141	CGGAGAAAAGCTGGTCGGTATTTCTGAGGATGAGTGGCATCTCAGGATGCACTTGGC	1200			
DB	1141	CGGAGAAAAGCTGGTCGGTATTTCTGAGGATGAGTGGCATCTCAGGATGCACTTGGC	1200			
QY	1201	CTGACTCGGGCAGCCGCTATGACATCATGAGCAATTCGGGGGACACTTCAGTCCGCGT	1260			
DB	1201	CTGACTCGGGCAGCCGCTATGACATCATGAGCAATTCGGGGGACACTTCAGTCCGCGT	1260			
QY	1261	GCATATCTGCTGGGATGTGGCAGATGCTTGAAGCGCTGAAAGCGGTGACCAAGGGA	1320			
DB	1261	GCATATCTGCTGGGATGTGGCAGATGCTTGAAGCGCTGAAAGCGGTGACCAAGGGA	1320			
QY	1321	TATAAGGACCCAGAAATGGGGCAGCTACATAAAGGATTTAGTGTGTCGGGCGAGGA	1380			
DB	1321	TATAAGGACCCAGAAATGGGGCAGCTACATAAAGGATTTAGTGTGTCGGGCGAGGA	1380			
QY	1381	AAAAAAATCTACTTGTGTAACCCCGACCTGAAGCGGAGAGAGCGTCACTATGAGGCT	1440			
DB	1381	AAAAAAATCTACTTGTGTAACCCCGACCTGAAGCGGAGAGAGCGTCACTATGAGGCT	1440			
QY	1441	GGGTGTATTACGATAAACCCTGGCTCTGAATGCCAATGTCAAGGTTTATGACTGAC	1500			
DB	1441	GGGTGTATTACGATAAACCCTGGCTCTGAATGCCAATGTCAAGGTTTATGACTGAC	1500			
QY	1501	TTCTCCAAACAGATTTGTCTTTATCCATAAATGATAACCAATAGCTATGAAACAGC	1560			
DB	1501	TTCTCCAAACAGATTTGTCTTTATCCATAAATGATAACCAATAGCTATGAAACAGC	1560			
QY	1561	GGAAAGCCCGGTTGACAGGTGTGGAATTTGCCGGCACATTCGCGTGTGGTCAAGGAT	1620			
DB	1561	GGAAAGCCCGGTTGACAGGTGTGGAATTTGCCGGCACATTCGCGTGTGGTCAAGGAT	1620			
QY	1621	GTACGCTGTCACTGAATTAACCTGCAGCCGAGTGAACAACTGATGATGATAACAA	1680			
DB	1621	GTACGCTGTCACTGAATTAACCTGCAGCCGAGTGAACAACTGATGATGATAACAA	1680			
QY	1681	GGTCCGCGCTGAGTTATACCCCTGAAACACATGGTGAATCGAAACTGAACTGGCAGATC	1740			
DB	1681	GGTCCGCGCTGAGTTATACCCCTGAAACACATGGTGAATCGAAACTGAACTGGCAGATC	1740			
QY	1741	ACCGAAGAGGTGGATCATGCTGGGTGCGGTTATCGGGGAAACACCACTTTTACC	1800			
DB	1741	ACCGAAGAGGTGGATCATGCTGGGTGCGGTTATCGGGGAAACACCACTTTTACC	1800			
QY	1801	CAGAAATTTATCGTCACTGAGCGTGTACAGAAAGTGTATGATGAGAAAGAGATATC	1860			
DB	1801	CAGAAATTTATCGTCACTGAGCGTGTACAGAAAGTGTATGATGAGAAAGAGATATC	1860			
QY	1861	CTGAAAGCCTGGACCGTGGTGGATGAGGTCTGTCTGGAGAGATGACGGATGCGCTGACG	1920			
DB	1861	CTGAAAGCCTGGACCGTGGTGGATGAGGTCTGTCTGGAGAGATGACGGATGCGCTGACG	1920			
QY	1921	CTGAAATGCTCGGTGAATAACCTGCTCAAGAGATTAACAGTACGCTGACCTGTACAGT	1980			
DB	1921	CTGAAATGCTCGGTGAATAACCTGCTCAAGAGATTAACAGTACGCTGACCTGTACAGT	1980			
QY	1981	GCCGGTGAAGAGTACGCTGTATGCGGTGATTAATCTCCAGACGGGATCATCAACACAGGA	2040			
DB	1981	GCCGGTGAAGAGTACGCTGTATGCGGTGATTAATCTCCAGACGGGATCATCAACACAGGA	2040			
QY	2041	TATGTGATACCTGAGCGAAATTTACTGATGTCGTGAACCTATCAGTTCTGA	2091			

/gene="t1pb"
/note="similar to antimicrobial-resistance protein;
similar to TerC of Alcaligenes sp.: SwissProt Accession
Number P18780"
/codon_start=1
/transl_table=11
/product="putative tellurium resistance protein D"
/protein_id="RAF36436.1"
/db_xref="GI:7108484"
/translation="NM_016181.1:1-1000
TDHIWHGYSIAATSLVVLGALGILASWMPFGTPESEKGR"

ORIGIN		Query Match	99.9%; Score 2089.4; DB 1; Length 8040;
		Best Local Similarity 100.0%; Pred. No. 0;	
		Matches 2090; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
QY	1	ATGCGAATAACCACTCTGCTTCCGTTAGTTCCTGTCGGAATTTTTCAGCCAGCAGC	60
Db	3036	ATGCGAATAACCACTCTGCTTCCGTTAGTTCCTGTCGGAATTTTTCAGCCAGCAGC	3095
QY	61	ATAGCTGCTGCAGAGATGATGATTTCTCGGCATCCGGCTATGAGAAAAGCTGACT	120
Db	3096	ATAGCTGCTGCAGAGATGATGATTTCTCGGCATCCGGCTATGAGAAAAGCTGACT	3155
QY	121	AACGACGCCAGTGTTCCTGATAGCCAGAGGAATTCAGTCCAGCCAGTACCAC	180
Db	3156	AACGACGCCAGTGTTCCTGATAGCCAGAGGAATTCAGTCCAGCCAGTACCAC	3215
QY	181	GATCTGGCGGAGCTCTGAGATCAGTAGAGGGTGTGGATTTGAAAGTGGTACGGGTAAA	240
Db	3216	GATCTGGCGGAGCTCTGAGATCAGTAGAGGGTGTGGATTTGAAAGTGGTACGGGTAAA	3275
QY	241	ACCGAGGGCTGGAATACGATCCGAGGAATCCAGCCAGTTCACCGTGTATCTGATT	300
Db	3276	ACCGAGGGCTGGAATACGATCCGAGGAATCCAGCCAGTTCACCGTGTATCTGATT	3335
QY	301	GATGGTTCCTCAGCGCGGAGCAGTGCAGTCTCCCAACGGTTTTTCTCCCATGAAT	360
Db	3336	GATGGTTCCTCAGCGCGGAGCAGTGCAGTCTCCCAACGGTTTTTCTCCCATGAAT	3395
QY	361	ACCGGGTTATGCTCCCTCTGCGCCATTTGAGCGTATGAGGGTTATCAGGGGGCGGATG	420
Db	3396	ACCGGGTTATGCTCCCTCTGCGCCATTTGAGCGTATGAGGGTTATCAGGGGGCGGATG	3455
QY	421	TCACACTGATGGCTCTGATGCGATGGCGGTGGTGAATATCAATTACCAAGGAAT	480
Db	3456	TCACACTGATGGCTCTGATGCGATGGCGGTGGTGAATATCAATTACCAAGGAAT	3515
QY	481	GCAGACAATGGCTCTTCCGTCATGAGGCTGAATCTGCAGGAAAGCAACAATGG	540
Db	3516	GCAGACAATGGCTCTTCCGTCATGAGGCTGAATCTGCAGGAAAGCAACAATGG	3575
QY	541	GGTAACAGCAGCCAGTTAATTTCTGGAGAGTGGTCCCTTGTGGATGTTCTGTCAGC	600
Db	3576	GGTAACAGCAGCCAGTTAATTTCTGGAGAGTGGTCCCTTGTGGATGTTCTGTCAGC	3635
QY	601	CTCAGGTACGGGTAGCACACAACAGCGTCAGGGTTTCATCGGTCACTCATCTGAGCGAT	660
Db	3636	CTCAGGTACGGGTAGCACACAACAGCGTCAGGGTTTCATCGGTCACTCATCTGAGCGAT	3695
QY	661	ACAGCAGCAGCGGTATTCCTTATCCACGGAGTCAAGAAATTAATCTTGGTGCAGCT	720
Db	3696	ACAGCAGCAGCGGTATTCCTTATCCACGGAGTCAAGAAATTAATCTTGGTGCAGCT	3755
QY	721	CTTGACTGGAAGCGTCCGAGAGGATGCTCTGCTTTCATATGATATGATATCCACCGGAG	780
Db	3756	CTTGACTGGAAGCGTCCGAGAGGATGCTCTGCTTTCATATGATATGATATCCACCGGAG	3815
QY	781	CGTTATGATTAACCGGGATGGGCACTGGGAGTCTGACGGGGGGATATGACCGGACCTTG	840
Db	3816	CGTTATGATTAACCGGGATGGGCACTGGGAGTCTGACGGGGGGATATGACCGGACCTTG	3875

QY	841	CGCTATGAGCGAAACAAAATTTTTCAGCTGGCTATGATCATCTTTTACCTTCGGAACATGG	900
Db	3876	CGCTATGAGCGAAACAAAATTTTTCAGCTGGCTATGATCATCTTTTACCTTCGGAACATGG	3935
QY	901	AAATCGTATCTGAACCTGGAAACGAGACAGAAAATAAAGGTCTGTCAGCTTGTACACAGTGA	960
Db	3936	AAATCGTATCTGAACCTGGAAACGAGACAGAAAATAAAGGTCTGTCAGCTTGTACACAGTGA	3995
QY	961	CTGAAGCGCGACAAAATGGGGGCTTCCGCTCAGCCCGGGAGCTTAAAGNAATCGAACCTTT	1020
Db	3996	CTGAAGCGCGACAAAATGGGGGCTTCCGCTCAGCCCGGGAGCTTAAAGNAATCGAACCTTT	4055
QY	1021	ATCCTGAATTCATTACTGCTTACCCCTCTGGGAGAAATCTCATCTGTTTACGTTGGGGGGC	1080
Db	4056	ATCCTGAATTCATTACTGCTTACCCCTCTGGGAGAAATCTCATCTGTTTACGTTGGGGGGC	4115
QY	1081	GAGTTTCAGAGCTCGTCCATGAAGACGAGAGTTGTCTCTCCAGCAGCAGGTGAACCTTTC	1140
Db	4116	GAGTTTCAGAGCTCGTCCATGAAGACGAGAGTTGTCTCTCCAGCAGCAGGTGAACCTTTC	4175
QY	1141	CGGAGAAAAGCTGGTCCGTTATTTGCTGAGGATGAGTGGCATCTTCACGATGCACTTGCG	1200
Db	4176	CGGAGAAAAGCTGGTCCGTTATTTGCTGAGGATGAGTGGCATCTTCACGATGCACTTGCG	4235
QY	1201	CTGACTCGGGGCGAGCCCTATGAACATCATAGCAATTCGGGGGACATTCAGTCCGCGT	1260
Db	4236	CTGACTCGGGGCGAGCCCTATGAACATCATAGCAATTCGGGGGACATTCAGTCCGCGT	4295
QY	1261	GCATATCTGCTGCTGGATGTCGAGATGCTTGAACGCTGAAGCGGTGTGACCCAGGGA	1320
Db	4296	GCATATCTGCTGCTGGATGTCGAGATGCTTGAACGCTGAAGCGGTGTGACCCAGGGA	4355
QY	1321	TATAAGGACACCAAGATGGGGGAGCTACATAAAGGGATTTAGTGTGTGTCTCGGSCAGGGA	1380
Db	4356	TATAAGGACACCAAGATGGGGGAGCTACATAAAGGGATTTAGTGTGTGTCTCGGSCAGGGA	4415
QY	1381	AAAAAATCTACTGTTGTAACCCCGACCTGAAGCGGAAAGAGCGCTCAGTTATGAGGCT	1440
Db	4416	AAAAAATCTACTGTTGTAACCCCGACCTGAAGCGGAAAGAGCGCTCAGTTATGAGGCT	4475
QY	1441	GGGTGTATTACGATTAACCCCGGCTGTAATGCCAATGTACAGGTTTATCAGCTGAC	1500
Db	4476	GGGTGTATTACGATTAACCCCGGCTGTAATGCCAATGTACAGGTTTATCAGCTGAC	4535
QY	1501	TTCTCCAAACAGATTGCTCTTATTCATAAATGATTAACCAACCAATAGCTATGTAACAGC	1560
Db	4536	TTCTCCAAACAGATTGCTCTTATTCATAAATGATTAACCAACCAATAGCTATGTAACAGC	4595
QY	1561	GGAAAGCCCGGTTGACGGTGTGGAATTTCCCGGACATTTCCGCTGTGTGTCAGAGAT	1620
Db	4596	GGAAAGCCCGGTTGACGGTGTGGAATTTCCCGGACATTTCCGCTGTGTGTCAGAGAT	4655
QY	1621	GTACCGCTGTCACTGAATTTACACCTGGACCCGAAAGTGAACCAACCTGATGTTGTAACAA	1680
Db	4656	GTACCGCTGTCACTGAATTTACACCTGGACCCGAAAGTGAACCAACCTGATGTTGTAACAA	4715
QY	1681	GGTGGCGGCTGATGTTATACCCCTGGAACACATGTTGTAATGCGAAACTGAACCTGAGATC	1740
Db	4716	GGTGGCGGCTGATGTTATACCCCTGGAACACATGTTGTAATGCGAAACTGAACCTGAGATC	4775
QY	1741	ACCGAAGAGTGGGATCATGCTGGTCCCGTTATCGCGGAAAAACACCACTTTTCACC	1800
Db	4776	ACCGAAGAGTGGGATCATGCTGGTCCCGTTATCGCGGAAAAACACCACTTTTCACC	4835
QY	1801	CAGAAATTTTCCTCATCTGAGCGCTGTACAGAAAGAAAGTGTATGATGAAAGGAGATATAC	1860
Db	4836	CAGAAATTTTCCTCATCTGAGCGCTGTACAGAAAGAAAGTGTATGATGAAAGGAGATATAC	4895
QY	1861	CTGAAGCCTGACCGTGGTGGATGCTGCTGCTGGAAGATGACGGATGCCCTGACG	1920
Db	4896	CTGAAGCCTGACCGTGGTGGATGCTGCTGCTGGAAGATGACGGATGCCCTGACG	4955
QY	1921	CTGAATGCTGGGTGAATAAACCTGCTCAACAGGATTTACAGTGACGCTGTGACAGT	1980


```

/function="orf; Adaptations, atypical conditions"
/notes="Residues 1 to 192 of 192 are 100.00 pct identical
to residues 1 to 192 of 192 from GenPept 118 :
gi|950682|gb|AA86850.1| [U59239] putative TerD [Serratia
marcescens]"
/codon_start=1
/transl_table=11
/product="putative phase inhibition, colicin resistance
and tellurite resistance protein"
/protein_id="AAG55729.1"
/db_xref="GI:12514500"
/translations="MSVSLSKGGNVLSKAAPSMNVGLGMDARSTDQDFDLSDAS
AFILASIKVSGDFIFNNLTSSDGSVTHDNRITGGDGDSDSLKIKLDAVPSEV
DKLIFVTHDAQRSGFQVSGSARIRLVNDNQNEVARYDLTSDASTETAMVFEGL
YRNGEKKFRAVGQVAGGLASVCAQYGINAS"
4167..4742
/gene="terE_2"
/notes="terE_2"
/notes="synonym: Z1615"
4167..4742
/gene="terE_2"
/function="orf; Adaptations, atypical conditions"
/notes="Residues 1 to 191 of 191 are 100.00 pct identical
to residues 1 to 191 of 191 from GenPept 118 :
gi|710848|gb|AAF3643.1|AF126104_3 (AF126104) putative
tellurium resistance protein B [Escherichia coli O157:H7]"
/codon_start=1
/transl_table=11
/product="putative phase inhibition, colicin resistance
and tellurite resistance protein"
/protein_id="AAG55730.1"
/db_xref="GI:12514501"
/translations="MAVSLVKGNGVSLTKEAPTMNVVGLGMDARVTDGQFDDLDAS
VFVAGDEKVLSDAHPFNNKTSPPGAVEHQDNRTGGDGDDEQVIXDITKVSADI
KELVAVIYDAARKQNGFMSVFNRYNNDNGTEIARFDSLSDASTETAMVFEGL
YRGTETWKFAVGQVAGGLASALASQHGYNV"
5164..5472
/gene="terF_2"
/notes="terF_2"
/notes="synonym: Z1616"
5164..5472
/gene="terF_2"
/function="orf; Adaptations, atypical conditions"
/notes="Residues 1 to 102 of 102 are 100.00 pct identical
to residues 1 to 102 of 102 from GenPept 118 :
gi|710848|gb|AAF3643.1|AF126104_2 (AF126104) putative
tellurium resistance protein A [Escherichia coli O157:H7]"
/codon_start=1
/transl_table=11
/product="partial putative phase inhibition, colicin
resistance and tellurite resistance protein"
/protein_id="AAG55731.1"
/db_xref="GI:12514502"
/translations="MNLQSGNLPLOQSTIRNLQVPAKSGFKGEPDTCLEFMLNAQKG
VSGDSDFIFNNLSPEGAVRLVTSQQASIEIALDRVPANVSKIAITVVIDGETIS"
"
complement (6064..8154)
/gene="Z1617"
complement (6064..8154)
/gene="Z1617"
/function="putative membrane; Cell envelop; Outer membrane
constituents"
/notes="Residues 1 to 696 of 696 are 99.85 pct identical to
residues 1 to 696 of 696 from GenPept 118 :
gi|7108480|gb|AAF36432.1|AF126104_1 (AF126104) adhesin
[Escherichia coli O157:H7]"
/codon_start=1
/transl_table=11
/product="putative receptor"
/protein_id="AAG55732.1"

```

Query Match 99.9%; Score 2089.4; DB 1; Length 10029;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2090; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 1 ATGCGAATAACCACTCTGGCTTCGGTAGTCAATCCCTGTCTCGATTTTCAGCCACAGC 60
Db ATGCGAATAACCACTCTGGCTTCGGTAGTCAATCCCTGTCTCGATTTTCAGCCACAGC 8095
QY 61 ATAGCTCTCTCAGAGGATGTGATGATTTCTCGGCATCCGGCTATCAGAAAAGCTGACT 120
Db ATAGCTCTCTCAGAGGATGTGATGATTTCTCGGCATCCGGCTATCAGAAAAGCTGACT 8035
QY 121 AACCGAGCCGCCAGTGTCTGTGATTAGCCAGAGGAATTCAGTCCAGCCAGTACCAAC 180
Db AACCGAGCCGCCAGTGTCTGTGATTAGCCAGAGGAATTCAGTCCAGCCAGTACCAAC 7975
QY 181 GATCTGCGCGAGGCTCTGAGATCAGTAGAGGGTGTGGATGTTCAAGAGTGGTACGGTAAA 240
Db GATCTGCGCGAGGCTCTGAGATCAGTAGAGGGTGTGGATGTTCAAGAGTGGTACGGTAAA 7915
QY 241 ACCGAGGGCTGGAATTCAGCATCCGAGGAATGCCAGGAGTTTACACGCTGATCTGATT 300
Db ACCGAGGGCTGGAATTCAGCATCCGAGGAATGCCAGGAGTTTACACGCTGATCTGATT 7855
QY 301 GATGGTCTCTCAGCGCGGAGCAGTACGTGACTCCCAACGGTTTTCTCCCATGAT 360
Db GATGGTCTCTCAGCGCGGAGCAGTACGTGACTCCCAACGGTTTTCTCCCATGAT 7795
QY 361 ACCGGTTCATGCCCTCTCTGGCCCATTCAGCGTATTGAGGTTATCAGGGGGCGGATG 420
Db ACCGGTTCATGCCCTCTCTGGCCCATTCAGCGTATTGAGGTTATCAGGGGGCGGATG 7735
QY 421 TCACACTGTATGGCTCTGATCGATGGGGGTGTGGTGAATATCATATACCAAGAGAT 480
Db TCACACTGTATGGCTCTGATCGATGGGGGTGTGGTGAATATCATATACCAAGAGAT 7675
QY 481 GCAGACAAATGGCTCTCTTCCTCCGTCATCAGAGGCTGAAATCTCAGGAAAGCAAAATGG 540
Db GCAGACAAATGGCTCTCTTCCTCCGTCATCAGAGGCTGAAATCTCAGGAAAGCAAAATGG 7615
QY 541 GGTAAACAGCAGCAGTAAATTTCTGGAGCAGTGGTCCCTTTGTGGATGATTTCTCAGC 600
Db GGTAAACAGCAGCAGTAAATTTCTGGAGCAGTGGTCCCTTTGTGGATGATTTCTCAGC 7555
QY 601 CTCAGCTACGCGGTAGCACACAAACAGCGCTCAGGGTTCATCGGTACATCATCTGAGCGAT 660
Db CTCAGCTACGCGGTAGCACACAAACAGCGCTCAGGGTTCATCGGTACATCATCTGAGCGAT 7495
QY 661 ACAGCAGCAGCGGTATTCCTTATCCACGGAGTACAGAAATATAATCTTGGTGCAGCT 720
Db ACAGCAGCAGCGGTATTCCTTATCCACGGAGTACAGAAATATAATCTTGGTGCAGCT 7435
QY 721 CTTGACTGGAAGCGCTCGGAGCAGGATGTCTCTGTTTGTATATGATACCAACCCGGCAG 780
Db CTTGACTGGAAGCGCTCGGAGCAGGATGTCTCTGTTTGTATATGATACCAACCCGGCAG 7375
QY 781 CGTTATGATAACCGGATGGCAACTGGGGAGTCTGACGGGGGATATGACCGGACCCCTG 840
Db CGTTATGATAACCGGATGGCAACTGGGGAGTCTGACGGGGGATATGACCGGACCCCTG 7315
QY 841 CGCTATGAGCGAAACAAAATTTTCAGCTGGCTATGATCATCTTTCACCTTCGGAACATGG 900
Db CGCTATGAGCGAAACAAAATTTTCAGCTGGCTATGATCATCTTTCACCTTCGGAACATGG 7255
QY 901 AATCGTATCTGACTGGAAACGAGACAGAAATAAAGTCTGAGCTTGTACCGAGTGTGA 960
Db AATCGTATCTGACTGGAAACGAGACAGAAATAAAGTCTGAGCTTGTACCGAGTGTGA 7195
QY 961 CTGAAGCGGACAAAATGGGGGCTTTCGGGTACGCGCGGAGCTTAAGGAATCGAACCTT 1020
Db CTGAAGCGGACAAAATGGGGGCTTTCGGGTACGCGCGGAGCTTAAGGAATCGAACCTT 7135
QY 1021 ATCTGAATTCATCTGCTTACCCGTCTGGGAGAAATCTCATCTGCTTACGTTGGGGGGC 1080
Db ATCTGAATTCATCTGCTTACCCGTCTGGGAGAAATCTCATCTGCTTACGTTGGGGGGC 7075
QY 1081 GAGTTTCAGAGCTCGTCCATGAAACGAGTGTGTCTTCCAGCAGCAGGTTGAAACTTTC 1140

```


Db 7074 GAGTTTCAGAGCTGCTCATGAAGAGCGAGTTGCTCTGCCACGACGAGTGAACATTTTC 7015

QY 1141 CGGCAGAAAAGCTGGTTCGTTATTTCTGAGGATGAGTGGCATCTCAGGATGCACTTGGC 1200

Db 7014 CGGCAGAAAAGCTGGTTCGTTATTTCTGAGGATGAGTGGCATCTCAGGATGCACTTGGC 6955

QY 1201 CTGACTCGGCGAGCGCTATGAACATCATGAGCAATTCGGGGGACATCTCAGTCCGCGT 1260

Db 6954 CTGACTCGGCGAGCGCTATGAACATCATGAGCAATTCGGGGGACATCTCAGTCCGCGT 6995

QY 1261 GCATATCTGGTCTGGGATGCGCAGATGCTGGAAGCTGAAAGCGGTGTGACACAGGGA 1320

Db 6894 GCATATCTGGTCTGGGATGCGCAGATGCTGGAAGCTGAAAGCGGTGTGACACAGGGA 6935

QY 1321 TATAAGGCACCGAATGGGCGAGCTACATAAAGGGAATTAGTGGTGTGTCGGGCGAGGGA 1380

Db 6834 TATAAGGCACCGAATGGGCGAGCTACATAAAGGGAATTAGTGGTGTGTCGGGCGAGGGA 6775

QY 1381 AAAACAATCTACTTGTAAACCCGACCTGAAGCCGGAAGAGCGCTCAGTTATGAGGCT 1440

Db 6774 AAAACAATCTACTTGTAAACCCGACCTGAAGCCGGAAGAGCGCTCAGTTATGAGGCT 6715

QY 1441 GGGGTGTATTACGATACCCCGCGGTCTGAATGCCAATGTACAGGTTTATGACTGAC 1500

Db 6714 GGGGTGTATTACGATACCCCGCGGTCTGAATGCCAATGTACAGGTTTATGACTGAC 6655

QY 1501 TTCTCCACAGATGCTCTTATTCATTAATGAATGAACACCAATAGCTATGTAACAGC 1560

Db 6654 TTCTCCACAGATGCTCTTATTCATTAATGAATGAACACCAATAGCTATGTAACAGC 6595

QY 1561 GGAAGCCCGGTGTGACCGGTGGAATTTCCCGGACATTTGCGGTGCTGCTGAGAGAT 1620

Db 6594 GGAAGCCCGGTGTGACCGGTGGAATTTCCCGGACATTTGCGGTGCTGCTGAGAGAT 6535

QY 1621 GTACCGTGTCTACGATTAACCTGACCGGAGTGAACAAAGTGTGATGATGATGATGAT 1680

Db 6534 GTACCGTGTCTACGATTAACCTGACCGGAGTGAACAAAGTGTGATGATGATGATGAT 6475

QY 1681 GGTGCGCGCTGAGTTATACCTGTAACACATGCTGAATGCGAACTGGAACCTGGAAGATC 1740

Db 6474 GGTGCGCGCTGAGTTATACCTGTAACACATGCTGAATGCGAACTGGAACCTGGAAGATC 6415

QY 1741 ACCGAAGAGTGGCATCAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1800

Db 6414 ACCGAAGAGTGGCATCAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 6355

QY 1801 CAGAATTTATTCGTCACCTGAGCGCTGTACAGAGAAAGTGTATGATGAGAAAGGAGATAC 1860

Db 6354 CAGAATTTATTCGTCACCTGAGCGCTGTACAGAGAAAGTGTATGATGAGAAAGGAGATAC 6295

QY 1861 CTGAAGCCCTGGACCGTGTGATGAGGCTGTCTGCTGGAAGATGACGATGCTGCTGAGC 1920

Db 6294 CTGAAGCCCTGGACCGTGTGATGAGGCTGTCTGCTGGAAGATGACGATGCTGCTGAGC 6235

QY 1921 CTGAATGCTCGGTGATTAACCTGCTCAACAGGATTAAGTGAAGTGAAGTGAAGTGAAGT 1980

Db 6234 CTGAATGCTCGGTGATTAACCTGCTCAACAGGATTAAGTGAAGTGAAGTGAAGTGAAGT 6175

QY 1981 GCCGGTAAGAGTACGCTGTATGCGCGGTGATTCTCCAGACGGGATCATCAACACAGGA 2040

Db 6174 GCCGGTAAGAGTACGCTGTATGCGCGGTGATTCTCCAGACGGGATCATCAACACAGGA 6115

QY 2041 TATGTGATACCTGAGGAAATTAATCTGAGTGTGCTGAACTATCAGTTCTGA 2091

Db 6114 TATGTGATACCTGAGGAAATTAATCTGAGTGTGCTGAACTATCAGTTCTGA 6064

RESULT 5

AE005274/c 10155 bp DNA linear BCT 21-MAR-2001
 LOCUS Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 98
 DEFINITION of 155.

ACCESSION AE005274 AE005174
 VERSION AE005274.1 GI:12513984
 KEYWORDS
 SOURCE Escherichia coli O157:H7 EDL933
 ORGANISM Escherichia coli O157:H7 EDL933
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Escherichia.
 REFERENCE
 AUTHORS 1 (bases 1 to 10155)
 Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
 Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
 Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
 Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
 Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
 Welch,R.A. and Blattner,F.R.
 TITLE Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
 JOURNAL Nature 409 (6819), 529-533 (2001)
 MEDLINE 21074935
 PUBMED 11206551
 REFERENCE 2 (bases 1 to 10155)
 AUTHORS Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
 Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
 Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
 Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
 Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
 Welch,R.A. and Blattner,F.R.
 TITLE Direct Submission
 JOURNAL Submitted (22-OCT-2000) Laboratory of Genetics, University of
 Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
 FEATURES
 source
 1. 10155
 /organism="Escherichia coli O157:H7 EDL933"
 /mol_type="genomic DNA"
 /strain="EDL933"
 /serotype="O157:H7"
 /db_xref="taxon:155864"
 /note="enterohaemorrhagic"
 misc_feature
 <1..>10155
 /note="O-island #43; Region of the EDL933 chromosome not
 homologous to E. coli K-12 MG1655"
 gene
 242..550
 /gene="cerF"
 /note="synonym: Z1177"
 CDS
 242..550
 /gene="terF"
 /functions="orf; Adaptations, atypical conditions"
 /note="Residues 1 to 102 of 102 are 100.00 pct identical
 to residues 1 to 102 of 102 from GenPept 118;
 gi|710848|gb|AAF36433.1|AF126104_2 (AF126104) putative
 tellurium resistance protein A [Escherichia coli O157:H7]"
 /codon_start=1
 /transl_table=11
 /product="partial putative phage inhibition, colicin
 resistance and tellurite resistance protein"
 /protein_id="AGS5322.1"
 /db_xref="GI:12513985"
 /translation="MNLQSGNIPLOQSTIRLNLOYPAKSGFKGEPDTCFLFMLNAQGGK
 VGSDDSFIFYNNLSPEGAVRLVTGQQASIEIALDRVPANVSKIAITVVIDGETIS
 "
 complement (1142..3232)
 /gene="Z1178"
 complement (1142..3232)
 /genes="Z1178"
 /function="putative membrane; Cell envelop: Outer membrane
 constituents"
 /note="Residues 1 to 696 of 696 are 99.85 pct identical to
 residues 1 to 696 of 696 from GenPept 118;
 gi|7108480|gb|AAF36432.1|AF126104_1 (AF126104) adhesin
 [Escherichia coli O157:H7]"
 /codon_start=1
 /transl_table=11
 /product="putative receptor"
 /protein_id="AGS5323.1"
 /db_xref="GI:12513986"

/translation="MRITTLASVVIPLGFSASSIAAEDMIVSASGYEKKLTNA
SVVISELQSSDHLAALASVEGVDSVSGTGGLEISIRGMPASVYLLIDG
VROGSSDVTPNGSAMNTGMPPLAIEIRIVIRGPMSTLYGSDAMGQVNIIRKN
ADKWSLVNAGLQENKNGNSQNFWSGSLVDVDSVLSQVRGSGQORQSSVTSL
SDTAATRPETSCYNLCAKLDKASEQDLVDFMDTTRGQYDNDRDQSLGVTSL
DRTAEIRPNKISAGYDSTFTCTWKSVMWNETENGRILVRSVLKRDKXGLAGPRE
LKEKNLINSLLPLGESHVTVGSEFQSSNMKGTVLSTGTEIFRQKSWVFARDE
WHLDALATAGSYEHEQFGHFSFAYLVWDVADMTLGGVTTGYKAPRMQQLH
KISGVSQGNKLLGNPDLPKPEESVSYEAGVYDNPAGLNANVTGTFDFNKIVSY
SINDTNSVNSKARLHGFVFAFTPLNSEDVTLNVTWTRSEQRDGNKGLPSY
TPEHVNKLNMQITEEVASWLGARYGKTFRTQNYSSLSAVQKVKYDEKEVYLKAW
TVVDAGLSWMTDALTLNAAVNNLNKDYSDVLSYAGKSTLYAGDYFQTGSGSTTGYV
IPERNYWSLNYQF"
/gene 3429
/db_xref="GI:1251392"
/gene="Z1179"
/gene="Z1179"
/function="orf; Unknown function"
/note="No significant matches"
/codon_start=1
/transl_table=11
/protein_id="AAG55324.1"
/db_xref="GI:12513987"
/translation="MPVNSYPYGVGFANKKULL"
complement(3902..4189)
/gene="Z1180"
complement(3902..4189)
/gene="Z1180"
/function="orf; Unknown function"
/note="No significant matches"
/codon_start=1
/transl_table=11
/protein_id="AAG55325.1"
/db_xref="GI:12513988"
/translation="MSVINDRSMGKTPEFVDGNLFLRFLPHRYMRORFFOTHPMG
QAWFOEIRNTLGYCSLKSILCNSNPVXRGVTSILKIKVVMCIKPL"
4686..4904
/gene="Z1181"
4686..4904
/gene="Z1181"
/function="orf; Unknown function"
/note="Residues 1 to 58 of 72 are 41.37 pct identical to
residues 159 to 216 of 229 from GenPept 118:
gi|5880473|gb|AAD54666.1|AF097520.4 (AF097520) unknown
[Shigella flexneri]"
/codon_start=1
/transl_table=11
/protein_id="AAG55326.1"
/db_xref="GI:12513989"
/translation="MPKYAEMLMAEEDDKVCELIFKQDNLLKINVRPPIIKKYR
TLOCKVVMWILPVIEWLIVCTLVNS"
complement(5306..5404)
/gene="Z1182"
complement(5306..5404)
/gene="Z1182"
/function="orf; Unknown function"
/note="No significant matches"
/codon_start=1
/transl_table=11
/protein_id="AAG55327.1"
/db_xref="GI:12513990"
/translation="MIIITLIGIHILQRIIYNWYSNYLTKIQ"
5537..5872
/gene="Z1183"
5537..5872
/gene="Z1183"
/function="orf; Unknown function"
/note="Residues 68 to 111 of 111 are 50.00 pct identical
to residues 21 to 54 of 90 from GenPept 118:
gi|3661496|gb|AAC61726.1| (AF081285) neurotensin receptor
[Escherichia coli]"
/codon_start=1
/transl_table=11

/protein_id="AAG55328.1"
/db_xref="GI:12513991"
/translation="MVISVGYEKRGTGSSASSIIICKELQPDNAILIQKKAISVGRN
VEISTKRGSDINTRIPASNCLSARNALKHNTGAILMSYAFSIVTTNENKLAG
ANIKRCDTV"
6389..6523
/gene="Z1184"
6389..6523
/gene="Z1184"
/function="orf; Unknown function"
/note="No significant matches"
/codon_start=1
/transl_table=11
/protein_id="AAG55329.1"
/db_xref="GI:12513992"
/translation="MLQIMTRWISLDAYDMHHVNDYATECMTHIKALGSLITMSC
"
complement(6653..6847)
/gene="Z1185"
complement(6653..6847)
/gene="Z1185"
/function="orf; Unknown function"
/note="No significant matches"
/codon_start=1
/transl_table=11
/protein_id="AAG55330.1"
/db_xref="GI:12513993"
/translation="MKNKIKRLQLQARQREFATKNGRAALPHERLTYFPFLEDLIK
NNGKXRSKKHFLKEDK"
7167..7439
/gene="Z1186"
7167..7439
/gene="Z1186"
/function="orf; Unknown function"
/note="No significant matches"
/codon_start=1
/transl_table=11
/protein_id="AAG55331.1"
/db_xref="GI:12513994"
/translation="MSYQGLVLPDLRRGYLSKNASSNRFGLVNGDLSKDSRCW
CTFSRWRSSMDNAIACSSSLIVCTWEALYRSTSHYGNVLLQLA"
7831..7938
/gene="Z1187"
7831..7938
/gene="Z1187"
/function="orf; Unknown function"
/note="No significant matches"
/codon_start=1
/transl_table=11
/protein_id="AAG55332.1"
/db_xref="GI:12513995"
Query Match 99.9%; Score 2089.4; DB 1; Length 10155;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2090; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 ATCGAATAACCACTCTGGCTCCGAGTCACTTCCTGTCTCGGATTTTCACCCAGCAGC 60
Db 3232 ATCGAATAACCACTCTGGCTCCGAGTCACTTCCTGTCTCGGATTTTCACCCAGCAGC 3173
Qy 61 ATAGCTGCTGCAGAGGATGTGATGATGTCCTCGGATCCGGCTATGAGAAAAGCTGACT 120
Db 3172 ATAGCTGCTGCAGAGGATGTGATGATGTCCTCGGATCCGGCTATGAGAAAAGCTGACT 3113
Qy 121 AACGCAAGCCAGTGTCTTCTGTGATTAGCAGGAGGAATTCAGTCCAGCCAGTACCAC 180
Db 3112 AACGCAAGCCAGTGTCTTCTGTGATTAGCAGGAGGAATTCAGTCCAGCCAGTACCAC 3053
Qy 181 GATCTGCGGAGGCTCTGAGATCAGTACGAGGTGTGGATGTTGAAAGTGGTACGGGTAAA 240
Db 3052 GATCTGCGGAGGCTCTGAGATCAGTACGAGGTGTGGATGTTGAAAGTGGTACGGGTAAA 2993
Qy 241 ACCGAGGAGGCTGGAATCAGCATCCGAGGAATGCCAGCCAGTTCACCGCTGATCTGATT 300

Db 2992 ACCGAGGCTGGAAATCAGCATCCGAGGAATCCAGCCAGTTACACGCTGATCTGATT 2933
 QY 301 GATGGTGTTCCTAGGCGGAGCAGTGAACGTAAGTCCCAACGGTTTTTCTGCCATGAAT 360
 Db 2932 GATGGTGTTCCTAGGCGGAGCAGTGAACGTAAGTCCCAACGGTTTTTCTGCCATGAAT 2873
 QY 361 ACCGGGTTTCATGCCCCCTCTGCGCGCATTGAGCGTATTCAGGTTATCAGGGGCGCATG 420
 Db 2872 ACCGGGTTTCATGCCCCCTCTGCGCGCATTGAGCGTATTCAGGTTATCAGGGGCGCATG 2813
 QY 421 TCCACACTGTATGCTCTGATCGATGCGCGGTGTGGTGAATATCAFTTACCAGAAAGAAT 480
 Db 2812 TCCACACTGTATGCTCTGATCGATGCGCGGTGTGGTGAATATCAFTTACCAGAAAGAAT 2753
 QY 481 GCACACAAATGGCTCTCTCCGTCATGCGAGGCTGATCTGCAGGAAGCAACAATGG 540
 Db 2752 GCACACAAATGGCTCTCTCCGTCATGCGAGGCTGATCTGCAGGAAGCAACAATGG 2693
 QY 541 GGTAAACAGCAGCCAGTTTAAATTTCTGAGCAGTGGTCCCTTGTGGATGATCTGTCCAGC 600
 Db 2692 GGTAAACAGCAGCCAGTTTAAATTTCTGAGCAGTGGTCCCTTGTGGATGATCTGTCCAGC 2633
 QY 601 CTGACAGTACCGGTAGCAGACACACAGCGTCAGGGTTTCATCGGTCAATCATCTGAGCGAT 660
 Db 2632 CTGACAGTACCGGTAGCAGACACACAGCGTCAGGGTTTCATCGGTCAATCATCTGAGCGAT 2573
 QY 661 ACAGCAGGACCGCTATCTCTTATCCACAGCGTCAAGAAATATATCTTGGTGAACGT 720
 Db 2572 ACAGCAGGACCGCTATCTCTTATCCACAGCGTCAAGAAATATATCTTGGTGAACGT 2513
 QY 721 CTTGACAGGAGGCTCGAGCAGGATGTCTCTGTTTGAATGATACACACCGGCGAG 780
 Db 2512 CTTGACAGGAGGCTCGAGCAGGATGTCTCTGTTTGAATGATACACACCGGCGAG 2453
 QY 781 CGTTATGATACCGGATGCGCACTGGGAGTCTGACGGGGGATGATGACGGGCGCTG 840
 Db 2452 CGTTATGATACCGGATGCGCACTGGGAGTCTGACGGGGGATGATGACGGGCGCTG 2393
 QY 841 CGTTATGAGCAAAACAAATTTTCAGCTGGCTATGATCATCTTCACTTCCGAAACATGG 900
 Db 2392 CGTTATGAGCAAAACAAATTTTCAGCTGGCTATGATCATCTTTCACCTTCCGAAACATGG 2333
 QY 901 AATTCGTATCTGAATCGACAGCAGACAGAAATTAAGGTCTGAGCTTGTACGCTGTGA 960
 Db 2332 AATTCGTATCTGAATCGACAGCAGACAGAAATTAAGGTCTGAGCTTGTACGCTGTGA 2273
 QY 961 CTGAAGCGCACAATGCGGCTTGGCGGTTCAGCGTCCGCGGAGCTTAAAGAAATCGAACCTT 1020
 Db 2272 CTGAAGCGCACAATGCGGCTTGGCGGTTCAGCGTCCGCGGAGCTTAAAGAAATCGAACCTT 2213
 QY 1021 ATCCTGAATTCATCTGCTTACCTCTTGGGAGAAATCTCATCTGCTTACGCTGGGGGCG 1080
 Db 2212 ATCCTGAATTCATCTGCTTACCTCTTGGGAGAAATCTCATCTGCTTACGCTGGGGGCG 2153
 QY 1081 GAGTTTCAGAGCTCGCTCAATGAAGACGAGTGTCTCTGCGCAGCAGACGTTCAAACTTTC 1140
 Db 2152 GAGTTTCAGAGCTCGCTCAATGAAGACGAGTGTCTCTGCGCAGCAGACGTTCAAACTTTC 2093
 QY 1141 CGCAGAAAGAGTGTGCTGATTTGCTGAGGATGATGCGCATCTCACGGATGCACTTCCG 1200
 Db 2092 CGCAGAAAGAGTGTGCTGATTTGCTGAGGATGATGCGCATCTCACGGATGCACTTCCG 2033
 QY 1201 CTGACTCGGGCAGCGCTATGACATCATGACCAATTCGGGGGACCTTCACTGCTCGGCT 1260
 Db 2032 CTGACTCGGGCAGCGCTATGACATCATGACCAATTCGGGGGACCTTCACTGCTCGGCT 1973
 QY 1261 GCATATCTGCTGCGATGTCAGATGCTTGGACGCTGAAAGGCGGTGTGACCAACGGGA 1320
 Db 1972 GCATATCTGCTGCGATGTCAGATGCTTGGACGCTGAAAGGCGGTGTGACCAACGGGA 1913
 QY 1321 TATNAGGCACCAATGCGGCGCTTACATNAGGATGATGCTGCTGCTGCGGCGAGGA 1380

Db 1912 TATAAGGCACCCAGCAATGGGGCAGCTACTATAAGGATTAGTGTGTCTCCGGGAGGGA 1853
 QY 1381 AAAACAAATCTACTTGTGTAAACCCGACCTGAAGCGGGAAGAGAGCGTCACTTATGAGCT 1440
 Db 1852 AAAACAAATCTACTTGTGTAAACCCGACCTGAAGCGGGAAGAGAGCGTCACTTATGAGCT 1793
 QY 1441 GGGGTGTATTACGATAAACCAGCGCTCTGAATCCCAATGTCACAGGTTTATGACTGAC 1500
 Db 1792 GGGGTGTATTACGATAAACCAGCGCTCTGAATCCCAATGTCACAGGTTTATGACTGAC 1733
 QY 1501 TTCTCCAAACAAGATGTCTCTTATTTCCATAAATGATAACAACAATAGCTATGTAACAGC 1560
 Db 1732 TTCTCCAAACAAGATGTCTCTTATTTCCATAAATGATAACAACAATAGCTATGTAACAGC 1673
 QY 1561 GGAAGGCGCGGTTGCACGGTGTGGAATTTGCGCGCACATTTCCCGCTGTGTGTGAGGAT 1620
 Db 1672 GGAAGGCGCGGTTGCACGGTGTGGAATTTGCGCGCACATTTCCCGCTGTGTGTGAGGAT 1613
 QY 1621 GTACGCTGTCACTGAATTAACACCTGGAACCCGAAAGTGAACAACCTGATGGTGAACAAA 1680
 Db 1612 GTACGCTGTCACTGAATTAACACCTGGAACCCGAAAGTGAACAACCTGATGGTGAACAAA 1553
 QY 1681 GGTGCGCGCTGAGTATTAACCCCTGAAACACATGTGTAATGGAACCTGAACCTGACAGATC 1740
 Db 1552 GGTGCGCGCTGAGTATTAACCCCTGAAACACATGTGTAATGGAACCTGAACCTGACAGATC 1493
 QY 1741 ACCGAAGAGGTGGCATCATGCTGGGTGCGGTTATCGCGGAAACACACAGCTTTCACC 1800
 Db 1492 ACCGAAGAGGTGGCATCATGCTGGGTGCGGTTATCGCGGAAACACACAGCTTTCACC 1433
 QY 1801 CAGAATTAATTCCTCACTGAGCGCTGTACAGAAGAAAGTGTATGATGAGAAGGAGATAC 1860
 Db 1432 CAGAATTAATTCCTCACTGAGCGCTGTACAGAAGAAAGTGTATGATGAGAAGGAGATAC 1373
 QY 1861 CTGAAGCTGGAGCGGTGGATGACAGTCTGCTGCGGAAGATGACGGATGCGCTGACG 1920
 Db 1372 CTGAAGCTGGAGCGGTGGATGACAGTCTGCTGCGGAAGATGACGGATGCGCTGACG 1313
 QY 1921 CTGAATCTGCGGTGAATAACCTCTCAACAAGGATTTACAGTCAAGCTGAGCTGTACAGT 1980
 Db 1312 CTGAATCTGCGGTGAATAACCTCTCAACAAGGATTTACAGTCAAGCTGAGCTGTACAGT 1253
 QY 1981 GCCGTAAGAGTACCTGTATGCGCGGTGATTAATTCAGACGGGATCATCAACAACAGGA 2040
 Db 1252 GCCGTAAGAGTACCTGTATGCGCGGTGATTAATTCAGACGGGATCATCAACAACAGGA 1193
 QY 2041 TATGTGATCTGAGCGGAAATTAATGATGCTGCTGAACCTATCAGTTCTGA 2091
 Db 1192 TATGTGATCTGAGCGGAAATTAATGATGCTGCTGAACCTATCAGTTCTGA 1142

RESULT 6
 AR204161/c

LOCUS AR204161 87563 bp DNA linear PAT 20-JUN-2002
 DEFINITION Sequence 57 from patent US 6365723.
 ACCESSION AR204161
 VERSION AR204161.1 GI:21500738

KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 87563)
 AUTHORS Blattner, F.R., Burland, V., Plunkett, G. and Welch, R.

TITLE Sequences of E. coli O157
 JOURNAL Patent: US 6365723-A 57 02-APR-2002;

FEATURES
 source 1. .87563
 /organism="unknown"
 /mol_type="unassigned DNA"

Query Match 99.9%; Score 2089.4; DB 6; Length 87563;
 Best Local Similarity 100.0%; Pred. No. 0;

Matches 2090; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Qy	1	ATGCGAATAACCACTCTGGCTTCCTAGTCAATCCCTCTCGGATTTTCAGCAGCAGC	60						
Dd	49991	ATGCGAATAACCACTCTGGCTTCCTAGTCAATCCCTCTCGGATTTTCAGCAGCAGC	49932						
Qy	61	ATAGCTGCTGAGAGGATGTGATGATTTCTCGGCATCCGGCTATGAGAAAGCTGACT	120						
Dd	49931	ATAGCTGCTGAGAGGATGTGATGATTTCTCGGCATCCGGCTATGAGAAAGCTGACT	49872						
Qy	121	AACGAGCCGCCAGTGTCTGTGATAGCCAGGAGGAATTGCAGTCAGCCAGTACCAC	180						
Dd	49871	AACGAGCCGCCAGTGTCTGTGATAGCCAGGAGGAATTGCAGTCAGCCAGTACCAC	49812						
Qy	181	GATCTGGCGGAGCTCTGAGATCAGTAGAGGCTGTGATGTTGAAAGTGGTACCGGTAAA	240						
Dd	49811	GATCTGGCGGAGCTCTGAGATCAGTAGAGGCTGTGATGTTGAAAGTGGTACCGGTAAA	49752						
Qy	241	ACCGAGGCTCGAATCAGCATCCGAGGAATGCCAGGAGTTACACGCTGATCTGATT	300						
Dd	49751	ACCGAGGCTCGAATCAGCATCCGAGGAATGCCAGGAGTTACACGCTGATCTGATT	49692						
Qy	301	GATGCTGTCAGGCGGAGAGTGAAGTGAATGAGCGTATGAGGTTATCAGGGGCGGATG	360						
Dd	49691	GATGCTGTCAGGCGGAGAGTGAAGTGAATGAGCGTATGAGGTTATCAGGGGCGGATG	49632						
Qy	361	ACCGGTTTCATGCCCTCTGCGCCATTGAGCGTATGAGGTTATCAGGGGCGGATG	420						
Dd	49631	ACCGGTTTCATGCCCTCTGCGCCATTGAGCGTATGAGGTTATCAGGGGCGGATG	49572						
Qy	421	TCCACACTGATGGCTCTGATCGATGGCGGCTGTGGTGAATATCATTAACGAGAAAGAT	480						
Dd	49571	TCCACACTGATGGCTCTGATCGATGGCGGCTGTGGTGAATATCATTAACGAGAAAGAT	49512						
Qy	481	GCAGCAATAGCTCTCTTCCTGATGAGCGTGAATCTGCAGGAGCAACAAATGG	540						
Dd	49511	GCAGCAATAGCTCTCTTCCTGATGAGCGTGAATCTGCAGGAGCAACAAATGG	49452						
Qy	541	GSTAAACAGCAGCAGTTAATTTCTGGAGCAGTGGTCCCTTTGTGGATGATTTCTGCAGC	600						
Dd	49451	GSTAAACAGCAGCAGTTAATTTCTGGAGCAGTGGTCCCTTTGTGGATGATTTCTGCAGC	49392						
Qy	601	CTCAGAGTACCGGTAGCAGCAACAGCGTCAAGGTTTCATCGTCACTACCTGAGCGAT	660						
Dd	49391	CTCAGAGTACCGGTAGCAGCAACAGCGTCAAGGTTTCATCGTCACTACCTGAGCGAT	49332						
Qy	661	ACAGCAGCAGCGTATTCCTTATCCACGAGTCAAGAAATTAATTTCTGGTGCACGT	720						
Dd	49331	ACAGCAGCAGCGTATTCCTTATCCACGAGTCAAGAAATTAATTTCTGGTGCACGT	49272						
Qy	721	CTTGACTGGAAGCGTCCGAGCAGGATGTCTCTGGTTGATATGATACCCCGGAG	780						
Dd	49271	CTTGACTGGAAGCGTCCGAGCAGGATGTCTCTGGTTGATATGATACCCCGGAG	49212						
Qy	781	CGTTATGATAACCGGATGCGCACTGGGAGTCTGACGGGGGATATGACCGGACCTTG	840						
Dd	49211	CGTTATGATAACCGGATGCGCACTGGGAGTCTGACGGGGGATATGACCGGACCTTG	49152						
Qy	841	CGCTATCAGCGAACAATAATTTCTGCTGGCTATGATCATATTTTCCCTTCGGAACATGG	900						
Dd	49151	CGCTATCAGCGAACAATAATTTCTGCTGGCTATGATCATATTTTCCCTTCGGAACATGG	49092						
Qy	901	AAATCGTATCTGAACTGGAACGAGACAGAAATAAAGTCTGTAGCTGTACAGCTGTA	960						
Dd	49091	AAATCGTATCTGAACTGGAACGAGACAGAAATAAAGTCTGTAGCTGTACAGCTGTA	49032						
Qy	961	CTGAAGCGCACAAATGGGGCTTGGCGTCCGCTCAGCGCGGAGCTTAAGGAATCGAACCTT	1020						
Dd	49031	CTGAAGCGCACAAATGGGGCTTGGCGTCCGCTCAGCGCGGAGCTTAAGGAATCGAACCTT	48972						
Qy	1021	ATCCTGAATTCATTACTCTTACCCCTCTGGGAGAAATCTCATCTGGTTACGGTGGGGGC	1080						
Dd	48971	ATCCTGAATTCATTACTCTTACCCCTCTGGGAGAAATCTCATCTGGTTACGGTGGGGGC	48912						

RESULT 7

BD184766/c

LOCUS

86248 bp

DNA

linear

PAT 17-JUN-2003

Qy	1081	GAGTTTCAGAGCTCGTCCATGAAGACGGAGTTGTCTTCCAGCACAGGTGAATTC	1140
Dd	48911	GAGTTTCAGAGCTCGTCCATGAAGACGGAGTTGTCTTCCAGCACAGGTGAATTC	48852
Qy	1141	CGCAGAAAGAGCTGGTCCGTTATTTGTGAGGATGAGTGGCATCTCAGGATGACATTCG	1200
Dd	48851	CGCAGAAAGAGCTGGTCCGTTATTTGTGAGGATGAGTGGCATCTCAGGATGACATTCG	48792
Qy	1201	CTGACTGCGGGCAGCGCTATGAACATCATGAGCAATTCGGGGGACATTCAGTCCGCGT	1260
Dd	48791	CTGACTGCGGGCAGCGCTATGAACATCATGAGCAATTCGGGGGACATTCAGTCCGCGT	48732
Qy	1261	GCATATCTGCTCTGGGATGCGGATGCGTGAACGCTGAAGCGCGGTGTGACACGGGA	1320
Dd	48731	GCATATCTGCTCTGGGATGCGGATGCGTGAACGCTGAAGCGCGGTGTGACACGGGA	48672
Qy	1321	TATAAGCACCCAGATGCGGAGCTACATAAAGGATTAAGTGTGTCCGGGCGAGGGA	1380
Dd	48671	TATAAGCACCCAGATGCGGAGCTACATAAAGGATTAAGTGTGTCCGGGCGAGGGA	48612
Qy	1381	AAAAAATCTACTTGTGAACCCCGACCTGAAGCCGGAAGAGCGCTCACTTATGAGGCT	1440
Dd	48611	AAAAAATCTACTTGTGAACCCCGACCTGAAGCCGGAAGAGCGCTCACTTATGAGGCT	48552
Qy	1441	GGGTGTATTAAGATACCCCGCGCTCTGAATGCCAATGTCAAGTGTATGACTGAC	1500
Dd	48551	GGGTGTATTAAGATACCCCGCGCTCTGAATGCCAATGTCAAGTGTATGACTGAC	48492
Qy	1501	TTCTCCAAACAGATTTGTCTTATCCATAAATGATAACCAATAGCTATGTAACAGC	1560
Dd	48491	TTCTCCAAACAGATTTGTCTTATCCATAAATGATAACCAATAGCTATGTAACAGC	48432
Qy	1561	GGAAAGCCCGGTTGCAAGTGTGGAATTTCCGGGCACTTCCGCTGTGGTCAAGGAT	1620
Dd	48431	GGAAAGCCCGGTTGCAAGTGTGGAATTTCCGGGCACTTCCGCTGTGGTCAAGGAT	48372
Qy	1621	GTACGCTGTCACTGAATTAACCTGAGCCGGAAGTGAACAACTGATGTTGTAACAAA	1680
Dd	48371	GTACGCTGTCACTGAATTAACCTGAGCCGGAAGTGAACAACTGATGTTGTAACAAA	48312
Qy	1681	GCTGCGCGCTGAGTTATACCCCTGAACACATGATGTAATGCAAACTGCAATC	1740
Dd	48311	GCTGCGCGCTGAGTTATACCCCTGAACACATGATGTAATGCAAACTGCAATC	48252
Qy	1741	ACCGAAGGTGGCATCATGCTGGGTCCCGTTATCGCGGAAACACACGCTTTCACC	1800
Dd	48251	ACCGAAGGTGGCATCATGCTGGGTCCCGTTATCGCGGAAACACACGCTTTCACC	48192
Qy	1801	CAGAAATTTCTGCTCACTGAGCGCTGTACAGAAAGAAAGTGTATGATGAGAAAGGATAC	1860
Dd	48191	CAGAAATTTCTGCTCACTGAGCGCTGTACAGAAAGAAAGTGTATGATGAGAAAGGATAC	48132
Qy	1861	CTGAAAGCTGAGCGGTGGATGCGAGTCTGTCTGGAAGATGACGGATGCCCTGACG	1920
Dd	48131	CTGAAAGCTGAGCGGTGGATGCGAGTCTGTCTGGAAGATGACGGATGCCCTGACG	48072
Qy	1921	CTGATGCTGCGGTGAATAAATCTGCTCAACAGGATTAACGCTGAGCTGTACAGT	1980
Dd	48071	CTGATGCTGCGGTGAATAAATCTGCTCAACAGGATTAACGCTGAGCTGTACAGT	48012
Qy	1981	CGCGGTAAGAGTACGCTGTATGCGGCTGATTTACCTCCAGAGGATCATCAACACAGA	2040
Dd	48011	CGCGGTAAGAGTACGCTGTATGCGGCTGATTTACCTCCAGAGGATCATCAACACAGA	47952
Qy	2041	TATGTATACCTGAGCGAAATTTACTGGATGCTCGCTGAACACTATCAGTTCTGA	2091
Dd	47951	TATGTATACCTGAGCGAAATTTACTGGATGCTCGCTGAACACTATCAGTTCTGA	47901

DEFINITION	Nucleic acid molecule and polypeptide specific to intestinal hemorhagic pathogenic <i>Escherichia coli</i> O157:H7, and method of use thereof.	QY	361	ACGGGTTTCATGCCCCCTCTGGCCGCCCATTTGAGCGTTATCAGGGGCGCGATG	420
ACCESSION	BD184766	DB	49644	ACCGGTTTCATGCCCCCTCTGGCCGCCCATTTGAGCGTTATCAGGGGCGCGATG	49585
VERSION	JP 2002355074-A/2	QY	421	TCACACACTGTATGGCTCTGATGCGATGGCGGTGGTGAATATCATTACACAGAAAGAT	480
KEYWORDS	<i>Escherichia coli</i> O157:H7	DB	49584	TCACACACTGTATGGCTCTGATGCGATGGCGGTGGTGAATATCATTACACAGAAAGAT	49525
SOURCE	<i>Escherichia coli</i> O157:H7	QY	481	GCAGCAAAATGGCTCTTCCGTCAATCAGGGTGAATCTGCAGGAAGCAAAATGG	540
ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; <i>Escherichia</i> .	DB	49524	GCAGCAAAATGGCTCTTCCGTCAATCAGGGTGAATCTGCAGGAAGCAAAATGG	49465
REFERENCE	1 (Bases 1 to 86248)	QY	541	GCTAACACAGCAGCGATTAAATTTCTGGAGCAGTGTCCCTTGTGGATGATTTCTGCAGC	600
AUTHORS	Hayashi, H., Shinagawa, H., Makino, K., Hayashi, T., Onishi, S., Hattori, M. and Kurokawa, K.	DB	49464	GCTAACACAGCAGCGATTAAATTTCTGGAGCAGTGTCCCTTGTGGATGATTTCTGCAGC	49405
TITLE	Nucleic acid molecule and polypeptide specific to intestinal hemorhagic pathogenic <i>Escherichia coli</i> O157:H7, and method of use	QY	601	CTGCAGGTACGGGTAGCACACACAGCGTCCAGGGTTTCATCGTCAATCAGTACGCGAT	660
JOURNAL	Patent: JP 2002355074-A 2 10-DEC-2002;	DB	49404	CTGCAGGTACGGGTAGCACACACAGCGTCCAGGGTTTCATCGTCAATCAGTACGCGAT	49345
COMMENT	PRESIDENT OF UNIVERSITY OF TSUKUBA OS <i>Escherichia coli</i> O157:H7 PN JP 2002355074-A/2 PD 10-DEC-2002 PF 24-JAN-2002 JP 2002015959 PI HIDEO HAYASHI, HIDEO SHINAGAWA, KOZO MAKINO, TETSUYA HAYASHI, SHIN ONISHI,	QY	661	ACAGCAGCAGCGCGTATTCCTTTATCCACGAGTTCACAGAAATTAATCTTGTGTCAGGT	720
		DB	49344	ACAGCAGCAGCGCGTATTCCTTTATCCACGAGTTCACAGAAATTAATCTTGTGTCAGGT	49285
		QY	721	CTTGACTCGAAAGCGGTCCGAGCAGATGTGCTCTGGTTTGATATGGATACCCACCGGCGAG	780
		DB	49284	CTTGACTCGAAAGCGGTCCGAGCAGATGTGCTCTGGTTTGATATGGATACCCACCGGCGAG	49225
		QY	781	CGTTATGATAACCGGGATGGGCAACTGGGAGTCTGACGGGGGATATGACCGGACCGCT	840
		DB	49224	CGTTATGATAACCGGGATGGGCAACTGGGAGTCTGACGGGGGATATGACCGGACCGCT	49165
		QY	841	CGCTATGAGCAACAAATTTTCAGTGTGCTATCATATATCTTTCACCTTCGGAACATGG	900
		DB	49164	CGCTATGAGCAACAAATTTTCAGTGTGCTATCATATATCTTTCACCTTCGGAACATGG	49105
		QY	901	AAATCGTATCTGAACTCGAAGCAGACAGAAATAAAGTCTGTGAGCTTGTACGACGTGA	960
		DB	49104	AAATCGTATCTGAACTCGAAGCAGACAGAAATAAAGTCTGTGAGCTTGTACGACGTGA	49045
		QY	961	CTGAAGCGCGCAATGGCGCTTGCCTGCTGACCGCGGAGCTTAGGAAATCGAACCTT	1020
		DB	49044	CTGAAGCGCGCAATGGCGCTTGCCTGCTGACCGCGGAGCTTAGGAAATCGAACCTT	48985
		QY	1021	ATCCTGAATTCATTACTGCTTACCCCTCTGGGAGAACTCTCATCTGGTTACGTTGGGGGG	1080
		DB	48984	ATCCTGAATTCATTACTGCTTACCCCTCTGGGAGAACTCTCATCTGGTTACGTTGGGGGG	48925
		QY	1081	GAGTTTCAGAGCTCGTCCATGAAAGCAGGAGTTGCTCTGCCAGCAGAGTGAACCTTTC	1140
		DB	48924	GAGTTTCAGAGCTCGTCCATGAAAGCAGGAGTTGCTCTGCCAGCAGAGTGAACCTTTC	48865
		QY	1141	CGGCAAGAAAGCTGCTCGGTATTTGCTGAGCATGAGTGGCATCTCAGCGGATGCACCTTGC	1200
		DB	48864	CGGCAAGAAAGCTGCTCGGTATTTGCTGAGCATGAGTGGCATCTCAGCGGATGCACCTTGC	48805
		QY	1201	CTGACTCGGGCAGCCGCTATGAAATCATGAGCAATTCGGGGGACACTTTCAGTCCGCGT	1260
		DB	48804	CTGACTCGGGCAGCCGCTATGAAATCATGAGCAATTCGGGGGACACTTTCAGTCCGCGT	48745
		QY	1261	GCATATCTGCTGGGATGTGGCAGATGCTGACCGCTGAAAGCGGTGTGACACCGGA	1320
		DB	48744	GCATATCTGCTGGGATGTGGCAGATGCTGACCGCTGAAAGCGGTGTGACACCGGA	48685
		QY	1321	TATAAGCACCACCAATGGGGCAGCTACATAAAGGATTAGTGGTGTGTCGGGCGAGGA	1380
		DB	48684	TATAAGCACCACCAATGGGGCAGCTACATAAAGGATTAGTGGTGTGTCGGGCGAGGA	48625
		QY	1381	AAACAAATCTACTTGGTACCCCGACCTGAACCGGAGAGCGCTCAGTTATGAGGCT	1440
		DB	48624	AAACAAATCTACTTGGTACCCCGACCTGAACCGGAGAGAGCGCTCAGTTATGAGGCT	48565
		QY	1441	GGGGTGTATTACGATAACCCCGCGGTCTGAAATGCAATGTTCACAGGTTTATGACTGAC	1500
FEATURES	source				
ORIGIN					
Query Match	99.8%; Score 2087.8; DB 6; Length 86248;				
Best Local Similarity	99.9%; Pred. No. 0;				
Matches 2089; Conservative	0; Mismatches 2; Indels 0; Gaps 0;				
QY	1	ATCGAATACCACTCTGGCTCCGTAGTCAATCCCTGCTCGGATTTTCAGCCAGCAGC	60		
DB	50004	ATCGAATACCACTCTGGCTCCGTAGTCAATCCCTGCTCGGATTTTCAGCCAGCAGC	49945		
QY	61	ATAGTCTCTCAGAGGATGTGATGTCTCGCATCCGGTATGAGAAAGCTGACT	120		
DB	49944	ATAGTCTCTCAGAGGATGTGATGTCTCGCATCCGGTATGAGAAAGCTGACT	49895		
QY	121	AACGACCGCCAGTGTCTGTGATAGCAGGAGGAATTCAGTCCAGCAGTACCAAC	180		
DB	49884	AACGACCGCCAGTGTCTGTGATAGCAGGAGGAATTCAGTCCAGCAGTACCAAC	49825		
QY	181	GATCTCGCGAGGCTCTGAGATCAGTAGAGGGTGTGATGTTGAAAGTGGTACGGGTAA	240		
DB	49824	GATCTCGCGAGGCTCTGAGATCAGTAGAGGGTGTGATGTTGAAAGTGGTACGGGTAA	49765		
QY	241	ACCGGAGGCTGGAATTCAGATCCGAGAAATGCCAGCCAGTTACACGCTGATCTGATT	300		
DB	49764	ACCGGAGGCTGGAATTCAGATCCGAGAAATGCCAGCCAGTTACACGCTGATCTGATT	49705		
QY	301	GATGTTGTTCTCAGGGCGGAGCAGTACGCTGCCAACGGTTTTCTGCCATGAAT	360		
DB	49704	GATGTTGTTCTCAGGGCGGAGCAGTACGCTGCCAACGGTTTTCTGCCATGAAT	49645		

Db 282695 CTGCGAGTACGGGTAGCAACACACAGCGTCAAGGTTTATCGGTCAATCACTGACGAT 282636
QY 661 ACAGCAGGACCGGTATTCCTTTATCCACGAGTACAGAAATTAATCTTGGTCAAGT 720
Db 282635 ACAGCAGCCACGCGTATTCCTTTATCCACGAGTACAGAAATTAATCTTGGTCAAGT 282576
QY 721 CTTGACTGAAGCGCTGAGCAGAGATGCTCTGGTTTGTATGATGATACCAACCGGCGAG 780
Db 282575 CTTGACTGAAGCGCTGAGCAGAGATGCTCTGGTTTGTATGATGATACCAACCGGCGAG 282516
QY 781 CTTGATGATAACCGGATGGCAACTGGGGAGTCTGAACGGGGGATATGACCGGACCCCTG 840
Db 282515 CTTGATGATAACCGGATGGCAACTGGGGAGTCTGAACGGGGGATATGACCGGACCCCTG 282456
QY 841 CGCTATGACGGAAACAAATTTTCAGCTGCTATGATCATCTTTTACCTTCGACATGG 900
Db 282455 CGCTATGACGGAAACAAATTTTCAGCTGCTATGATCATCTTTTACCTTCGACATGG 282396
QY 901 AAATCGTATCTGAATCGGAACAGACAGAAATTAAGGTCGTGAGCTTGTACGAGTGA 960
Db 282395 AAATCGTATCTGAATCGGAACAGACAGAAATTAAGGTCGTGAGCTTGTACGAGTGA 282336
QY 961 CTGAACCGGACAAATGGGGGCTTGCCTGTCAGCCCGGGAGCTTAAGGAATCGAACCTT 1020
Db 282335 CTGAACCGGACAAATGGGGGCTTGCCTGTCAGCCCGGGAGCTTAAGGAATCGAACCTT 282276
QY 1021 ATCTGAATTCATCTGCTTACCTCTGCGGAGATCTCATCTGTTACGTTGGGGGC 1080
Db 282275 ATCTGAATTCATCTGCTTACCTCTGCGGAGATCTCATCTGTTACGTTGGGGGC 282216
QY 1081 GAGTTTCAGAGCTCGTCCATGAAGACGGAGTTGCTTCCAGCACAGGTGAACCTTTC 1140
Db 282215 GAGTTTCAGAGCTCGTCCATGAAGACGGAGTTGCTTCCAGCACAGGTGAACCTTTC 282156
QY 1141 CGGCAGAAAGCTGCTGCTATTTGCTGAGGATGAGTGCATCTCACGATCGACTGCG 1200
Db 282155 CGGCAGAAAGCTGCTGCTATTTGCTGAGGATGAGTGCATCTCACGATCGACTGCG 282096
QY 1201 CTGACTCGGGACCGCTATGAACATCATGAGCAATTCGGGGACACTTCAGTCCGGT 1260
Db 282095 CTGACTCGGGACCGCTATGAACATCATGAGCAATTCGGGGACACTTCAGTCCGGT 282036
QY 1261 GCATATCTGCTGAGGATGAGGAGATGCTGAGCGCTGAAGCGGCTGACACGGGA 1320
Db 282035 GCATATCTGCTGAGGATGAGGAGATGCTGAGCGCTGAAGCGGCTGACACGGGA 281976
QY 1321 TATAAGCCACCCAGATGGGCGAGCTACATAAAGGATTAGTGTCTGTCGGCGCAGGGA 1380
Db 281975 TATAAGCCACCCAGATGGGCGAGCTACATAAAGGATTAGTGTCTGTCGGCGCAGGGA 281916
QY 1381 AAAACAAATCTACTTGTAAACCCGACCTGAAGCCGGAAGAGCGCTGATGAGGCT 1440
Db 281915 AAAACAAATCTACTTGTAAACCCGACCTGAAGCCGGAAGAGCGCTGATGAGGCT 281856
QY 1441 GGGGTGATTAACATTAACCCCGCGTCTGAATGCCAATGTCAAGGTTTATGACTGAC 1500
Db 281855 GGGGTGATTAACATTAACCCCGCGTCTGAATGCCAATGTCAAGGTTTATGACTGAC 281796
QY 1501 TTCTCCAAAGATTTGCTCTTATTCATAAATGATAACCAATAGCTATGTAAACAGC 1560
Db 281795 TTCTCCAAAGATTTGCTCTTATTCATAAATGATAACCAATAGCTATGTAAACAGC 281736
QY 1561 GGAAGCCCGGTTGACCGGTGGAAATTTGCGGACATTTGCGGCTGTTGTCAGAGGAT 1620
Db 281735 GGAAGCCCGGTTGACCGGTGGAAATTTGCGGACATTTGCGGCTGTTGTCAGAGGAT 281676
QY 1621 GTCACGCTGCTACTGAATTAACCTGACCGGAGTGAACACGTCATGTGTATACAAA 1680
Db 281675 GTCACGCTGCTACTGAATTAACCTGACCGGAGTGAACACGTCATGTGTATACAAA 281616
QY 1681 GGTGCGCGCTGAGTTATACCCCTGAAACACATCGTGAATGCGAACTGAACTGGCAGATC 1740

Db 281615 GGTGCGCGCTGAGTTATACCCCTGAACACATGGTGAATCGAAACTGACGATC 281556
QY 1741 ACCGAAGAGTGGCAATCATGGTGGGTCCCGTATTCGCGGGAACACACCGTTTCACC 1800
Db 281555 ACCGAAGAGTGGCAATCATGGTGGGTCCCGTATTCGCGGGAACACACCGTTTCACC 281496
QY 1801 CAGAAATTAATTCGTCACCTGAGCGCTGTACAGAAGAAAGTGTATGATGAGAAAGGAGATAC 1860
Db 281495 CAGAAATTAATTCGTCACCTGAGCGCTGTACAGAAGAAAGTGTATGATGAGAAAGGAGATAC 281436
QY 1861 CTGAACCCCTGAGCGGTGGATGAGGTCCTGCTGGAAGATGACGGATGCCCTGACG 1920
Db 281435 CTGAACCCCTGAGCGGTGGATGAGGTCCTGCTGGAAGATGACGGATGCCCTGACG 281376
QY 1921 CTGAATGCTCGCGTGAATAACCTGCTCAACAAGAAATTTACAGTACGCTGACGCTGTACAGT 1980
Db 281375 CTGAATGCTCGCGTGAATAACCTGCTCAACAAGAAATTTACAGTACGCTGACGCTGTACAGT 281316
QY 1981 GCCCGTGAAGTACGCTGTATGCCGCTGATCTTCCAGACGGGATCATCAACACAGGA 2040
Db 281315 GCCCGTGAAGTACGCTGTATGCCGCTGATCTTCCAGACGGGATCATCAACACAGGA 281256
QY 2041 TATGTGATACCTGAGCGAAATTTACTGATGTCGCTGAACCTATCAGTTCTGA 2091
Db 281255 TATGTGATACCTGAGCGAAATTTACTGATGTCGCTGAACCTATCAGTTCTGA 281205

RESULT 9
AX702523/c 76804 bp DNA linear PAT 03-APR-2003
LOCUS AX702523 Sequence 109 from Patent WO02059320.
DEFINITION AX702523
ACCESSION AX702523
VERSION AX702523.1 GI:29537720
KEYWORDS
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.

REFERENCE 1
AUTHORS Blattner, F.R., Welch, R.A. and Burland, V.D.
TITLE DNA sequences of Escherichia coli CPT073
PATENT: WO 02059320-A 109 01-AUG-2002;
JOURNAL WISCONSIN ALUMNI RESEARCH FOUNDATION (US)

FEATURES
source location/Qualifiers
1..76804
/organism="Escherichia coli"
/mol_type="unassigned DNA"
/db_xref="taxon:562"
misc_feature 39751 /note="Unsure"
misc_feature 49372 /note="Unsure"
misc_feature 66539 /note="Unsure"
misc_feature 73693 /note="Unsure"
misc_feature 73705 /note="Unsure"
misc_feature 73737 /note="Unsure"
misc_feature 73759 /note="Unsure"
misc_feature 73761 /note="Unsure"
misc_feature 73786 /note="Unsure"
misc_feature 73774 /note="Unsure"
misc_feature 73781 /note="Unsure"
misc_feature 73798 /note="Unsure"
misc_feature 73802 /note="Unsure"

ORIGIN	/note="Unsure"									
Query Match	99.5%;	Score 2079.8;	DB 6;	Length 76804;						
Best Local Similarity	99.7%;	Pred. No. 0;								
Matches 2084;	Conservative	0;	Mismatches	7;	Indels	0;	Gaps	0;		
Qy	1	ATCGAATAACCACTCTGGCTCCGTAGTCATCCCTCTCCGATTTTCAGCCAGC	60							
Db	44573	ATCGAATAACCACTCTGGCTCCGTAGTCATCCCTCTCCGATTTTCAGCCAGC	44514							
Qy	61	ATAGCTGCTGCAGAGAGATGTGATGATGTCTCGGCATCCGGCTATGAGAAAAGCTGACT	120							
Db	44513	ATAGCTGCTGCAGAGAGATGTGATGATGTCTCGGCATCCGGCTATGAGAAAAGCTGACT	44454							
Qy	121	AACGACGCCCGCAGTGTTTCTGTGATTAGCCAGGAGGAATTGCAGTCCAGCCAGTACCAC	180							
Db	44453	AACGACGCCCGCAGTGTTTCTGTGATTAGCCAGGAGGAATTGCAGTCCAGCCAGTACCAC	44394							
Qy	181	GATCTGGCGGAGCTCTGAGATCAGTACAGGGTGTGAGTGTGAAGTGTACCGGTAAA	240							
Db	44393	GATCTGGCGGAGCTCTGAGATCAGTACAGGGTGTGAGTGTGAAGTGTACCGGTAAA	44334							
Qy	241	ACCGGAGGCTCGGAATCAGCATCCGAGGAATGCCAGCAGTTACACGCTGATCTGATT	300							
Db	44333	ACCGGTGGCTGGAATCAGCATCCGAGGAATGCCAGCAGTTACACGCTGATCTGATT	44274							
Qy	301	GATGGTGTCTCAGGGCGAAGAGTGAAGTGAATCCCAACGGTTTTTCTGCCATGAAT	360							
Db	44273	GATGGTGTCTCAGGGCGAAGAGTGAAGTGAATCCCAACGGTTTTTCTGCCATGAAT	44214							
Qy	361	ACCGGTTTCATCCCTCTGCGCCCATTTGAGCGTATTGAGGTTATCAGGGGCGGATG	420							
Db	44213	ACCGGTTTCATCCCTCTGCGCCCATTTGAGCGTATTGAGGTTATCAGGGGCGGATG	44154							
Qy	421	TCACACTGATGGCTCTGATCGGATGGCGGCTGTGGTGAATATCATTTACAGAAAAGAT	480							
Db	44153	TCACACTGATGGCTCTGATCGGATGGCGGCTGTGGTGAATATCATTTACAGAAAAGAT	44094							
Qy	481	GCAGCAATAGCTCTCTTCGTCAATCGAGGCTGATCTGCAGGAAGCAACAATGG	540							
Db	44093	GCAGCAATAGCTCTCTTCGTCAATCGAGGCTGATCTGCAGGAAGCAACAATGG	44034							
Qy	541	GGTAAACAGCAGCAGTTAAATTTCTGAGCAGTGGTCCCTTGTGGATGATTTCTGCAGC	600							
Db	44033	GGTAAACAGCAGCAGTTAAATTTCTGAGCAGTGGTCCCTTGTGGATGATTTCTGCAGC	43974							
Qy	601	CTGCAGGTACGGGTAGCACACACAGCGTCAAGGTTTCATCGTCACTCAGCGGAT	660							
Db	43973	CTGCAGGTACGGGTAGCACACACAGCGTCAAGGTTTCATCGTCACTCAGCGGAT	43914							
Qy	661	ACAGCAGGACCGCTATTTCCTTTATCCACGAGTCAAGAAATATATCTTTGGTGACGT	720							
Db	43913	ACAGCAGGACCGCTATTTCCTTTATCCACGAGTCAAGAAATATATCTTTGGTGACGT	43854							
Qy	721	CTTGACTGAAGCGGTGCGAGCAGGATGTCTGTGGTTTGAATPGGATACACCCGGCAG	780							
Db	43853	CTTGACTGAAGCGGTGCGAGCAGGATGTCTGTGGTTTGAATPGGATACACCCGGCAG	43794							
Qy	781	CGTTATGATAACCGGGATGGGCACTGGGGAGTCTGACGGGGGATATGACCGGACCTTG	840							
Db	43793	CGTTATGATAACCGGGATGGGCACTGGGGAGTCTGACGGGGGATATGACCGGACCTTG	43734							
Qy	841	CGCTATGAGCGAAAACAAATTTTCAGCTGGCTATGATCATATCTTTTCACCTTCGAAACATGG	900							
Db	43733	CGCTATGAGCGAAAACAAATTTTCAGCTGGCTATGATCATATCTTTTCACCTTCGAAACATGG	43674							
Qy	901	AAATCGTATCTGAACTGGAACGAGCAGAAATAAAGTCTGTGAGCTGTACCGATGTA	960							
Db	43673	AAATCGTATCTGAACTGGAACGAGCAGAAATAAAGTCTGTGAGCTGTACCGATGTA	43614							
Qy	961	CTGAAGCGCGAATAATGGGGCTTGCCGGTTCAGCCGGGAGCTTTAAGGAATCGAACCTT	1020							
Db	42533	TATGTGATACCTGAGCGAAATTAATGGAATGCTGCTGAACTATTCAGTTCTGA	2091							
Qy	2041	TATGTGATACCTGAGCGAAATTAATGGAATGCTGCTGAACTATTCAGTTCTGA	2091							
Db	42533	TATGTGATACCTGAGCGAAATTAATGGAATGCTGCTGAACTATTCAGTTCTGA	42483							


```
EDLRAHKFGVAGFIVDEIQITLTARFLILOAVEFIORTOCFQOQTQLGTVDRQPA
GGFGEFTFRADIEABVPFFELLVYVNAFQIVTGNVQLNATANVLQGNREBFTHTTC
HHAASNGDFDQRFQFVLFTFFVFCVQLVRGMATEIVREGNPLLTQVRLQLLTRFQF
IVKIDYRSALCVLFRHVGSS"
complement(4833..5931)
/genes="prfB"
/locus_tag="c5622"
complement(join(4833..5855,5857..5931))
/genes="prfB"
/locus_tag="c5622"
/function="factor; Proteins - translation and
modification"
/notes="RF-2; Escherichia coli K-12 ortholog: b2891;
Escherichia coli O157:H7 ortholog: z4229; ribosomal
slippage: an in-frame UGA termination codon is located
within the prfB sequence, and a naturally occurring +1
frameshift is required for synthesis of RF-2"
/codon_start=1
/transl_table=11
/product="Peptide chain release factor 2"
/protein_id="AAN8191.1"
/db_xref="GI:26109714"
/translation="MFEINPVNRRQDLTFRSDVLRGVLDYDAKERLEEYNABLEQ
DVNPERACALAKERSSEUAVDTLDQMKGLDVSGLLELAADDETFNEKVAE
LDALEKALCLFRFMFSGEYSDCYLDIQSGGTBAQWASMLERMYLRWASRG
FATEIEBEGEVAGIKSVTIKISGDIAYGNLTETGVRLVRKSPFDSGRRHTFS
SAFVPEVDDIDIENPADLRIDVVRASGAGQHVNRITESAVRITHPTGTQCN
DRSOHNKQAKOMKAKYLELMOKNAEKOMEDKSDIGWSQIRSYVLDSDRIK
DLRTGVTENTQVLDGSDQIEASLKAGL"
complement(6022..7755)
/genes="recJ"
/locus_tag="c3472"
complement(6022..7755)
/genes="recJ"
/locus_tag="c3472"
/EC_number="3.1.-.-"
/function="enzyme; Degradation of DNA"
/notes="Escherichia coli K-12 ortholog: b2892; Escherichia
coli O157:H7 ortholog: z4230"
/codon_start=1
/transl_table=11
/product="Single-stranded-DNA-specific exonuclease recJ"
/protein_id="AAN8192.1"
/db_xref="GI:26109715"
/translation="MKQIQILRRREVDETDLPABELPLRLRYASRGVRSQAHLERS
VKGMLQQLSGVEKAVELYNAREGRTIIVVDFDADGATSTALSVMRSLGCSN
IDYLVNRPEDGVSPEVVDQAHARGAQLIVTVNDG:SSHAGVEHARSLSGPIVITD
HILPGDTLPAABAIINPNLRDCNFPKSLAGVGAFYLMALRTFLRQGWPFDERGIA
IPNALLEDLVALGTVDVVDANNRILTWQMSRIIRAGKCRPGIKALLEVANDRPIQ
KLAASDLGALGPLRNAAGRLDDMSGVALLLCDNIGEARVLANSLDALNTRKRIEO
GMQVEALTCEKLRSDTLPGLAMYPHEWQGVVGLASIKERFHRPIAIPAPAG
DGTGLSGRSIOGLHNRDALERLDLYPGMLKPGHAWAAGLSDEKFLQFQRFQ
ELVTEWLDPSLQGVSVSDGPLSAEMETMEVALRDLADGPWGQMPPELFFGHFLLQ
QLRVGRHLKVMVEPVGGGFLLDGAFNVDTALMPDNGVREVQLAYKLIDINEPRGNS
LQIIIDNIWPI"
complement(7761..8471)
/genes="dsbC"
/locus_tag="c3473"
Query Match 99.5%; Score 2079.8; DB 1; Length 303121;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 2084; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 1 ATCGGAATACCACTCGGTCCTCGTAGTCATCCCTGCTCGGATTTTCACGCCAGCAGC 60
DB 123542 ATCGGAATACCACTCGGTCCTCGTAGTCATCCCTGCTCGGATTTTCACGCCAGCAGC 123483
QY 61 ATAGCTGCTCAGAGGATGTGATGATTTGTCGGCATCCGGTATCGAGAAAAGCTGACT 120
DB 123482 ATAGCTGCTCAGAGGATGTGATGATTTGTCGGCATCCGGTATCGAGAAAAGCTGACT 123423
QY 121 AAGCAGCCCGCAGTGTCTGTGATAGCAGGAGGATTCGAGTCCAGCCAGTACCAC 180
```

```
DB 123422 AACCGACGCCCAGTGTCTGTGATTTAGCCAGGAGGAATTGCAGTCCAGCCAGTACCAC 123363
QY 181 GATCTGGCGGAGGCTCTGAGATCAGTAGAGGGTGTGGATGTTCAAAGTGTGATCGGTAAA 240
DB 123362 GATCTGGCGGAGGCTCTGAGATCAGTAGAGGGTGTGGATGTTCAAAGTGTGATCGGTAAA 123303
QY 241 ACCGGAGGGCTGGAAATCAGCATCCGAGGAATCCAGCCAGTTCACACGCTGATCTGAT 300
DB 123302 ACCGGTGGCTGGAAATCAGCATCCGAGGAATCCAGCCAGTTCACACGCTGATCTGAT 123243
QY 301 GATGGTGTCTGTCAGCGCGGAAGCAGTGCAGCTCCCAACGGTCTTCTGCGCATGAAT 360
DB 123242 GATGGTGTCTGTCAGCGCGGAAGCAGTGCAGCTCCCAACGGTCTTCTGCGCATGAAT 123183
QY 361 ACCGGGTTCATGCCCTCTGCGCGGCATTTAGCGGTATTGAGGTTCATCAGGGGCGCGATG 420
DB 123182 ACCGGGTTCATGCCCTCTGCGCGGCATTTAGCGGTATTGAGGTTCATCAGGGGCGCGATG 123123
QY 421 TCACACTGTATGGCTCTGATGCGATGGCGGTGTGGTGAATATCATTTACCAGAAAGAT 480
DB 123122 TCACACTGTATGGCTCTGATGCGATGGCGGTGTGGTGAATATCATTTACCAGAAAGAT 123063
QY 481 GCAGACAAATGGCTCTCTCCGTCAATGCAGGGCTGAATCTGCAGGAAAGCAACAATGG 540
DB 123062 GCAGACAAATGGCTCTCTCCGTCAATGCAGGGCTGAATCTGCAGGAAAGCAACAATGG 123003
QY 541 GGTAACAGCAGCCAGTTTAAATTTCTGGAGCAGTGGTCCCTTGTGGATGATTTCTGTGACG 600
DB 123002 GGTAACAGCAGCCAGTTTAAATTTCTGGAGCAGTGGTCCCTTGTGGATGATTTCTGTGACG 122943
QY 601 CTGCAGGTACCGGTAGCACACACACGCTCAGGTTTCATCGTCAATCAGTACATGACGAT 660
DB 122942 CTGCAGGTACCGGTAGCACACACACGCTCAGGTTTCATCGTCAATCAGTACATGACGAT 122883
QY 661 ACAGCAGCAGCGGTATTTCTTATCCACGGAGTTCAGAGATTAATTAATCTTGTGACCGT 720
DB 122882 ACAGCAGCAGCGGTATTTCTTATCCACGGAGTTCAGAGATTAATTAATCTTGTGACCGT 122823
QY 721 CTTGACTGGAAGCGTGGAGCAGGATGTCTCTGTTGATGATGATACCCCGGCGAG 780
DB 122822 CTTGACTGGAAGCGTGGAGCAGGATGTCTCTGTTGATGATGATACCCCGGCGAG 122763
QY 781 CGTTATCATTAACCGGATGGGCAACTGGGAGTCTGACGGGGGATATGACCGGACCCCTG 840
DB 122762 CGTTATCATTAACCGGATGGGCAACTGGGAGTCTGACGGGGGATATGACCGGACCCCTG 122703
QY 841 CGCTATCAGCGAAACAAAAATTTCAGCTGGCTATGATCATCTTTACCTTCGGAAACATGG 900
DB 122702 CGCTATCAGCGAAACAAAAATTTCAGCTGGCTATGATCATCTTTACCTTCGGAAACATGG 122643
QY 901 AAATCGGTATCTGAACTGGAACGAGACAGAAAAATAAGGTCTGTGAGCTTGTACCGAGTGA 960
DB 122642 AAATCGGTATCTGAACTGGAACGAGACAGAAAAATAAGGTCTGTGAGCTTGTACCGAGTGA 122583
QY 961 CTGAAGCGCGACAAATGGGGGCTTGGCGGTACGCGGAGCTTAAGGAATCGAACCTT 1020
DB 122582 CTGAAGCGCGACAAATGGGGGCTTGGCGGTACGCGGAGCTTAAGGAATCGAACCTT 122523
QY 1021 ATCTGTAATTCATTACTGCTTACCCCTCTGGGAGATCTCATCTGTTACGTTGGGGGGC 1080
DB 122522 ATCTGTAATTCATTACTGCTTACCCCTCTGGGAGATCTCATCTGTTACGTTGGGGGGC 122463
QY 1081 GAGTTTCAGAGCTCGTCCATGAAAGACGGAGTGTCTCTGCCAGCAGAGTGAACCTTTC 1140
DB 122462 GAGTTTCAGAGCTCGTCCATGAAAGACGGAGTGTCTCTGCCAGCAGAGTGAACCTTTC 122403
QY 1141 CGGAGAAAAAGCTGCTCGTATTTCGTGAGGATGAGTGGCATCTCAGGATGACCTTGG 1200
DB 122402 CGGAGAAAAAGCTGCTCGTATTTCGTGAGGATGAGTGGCATCTCAGGATGACCTTGG 122343
QY 1201 CTGACTCGGGGACGCCCTATGAACATCATGAGCAATTCGGGGGACACTTTCAGTCCGCGT 1260
DB 122342 CTGACTCGGGGACGCCCTATGAACATCATGAGCAATTCGGGGGACACTTTCAGTCCGCGT 122283
```



```

/note="I8629"
/codon_start=1
/transl_table=11
/product="transposase R14"
/protein_id="AAC61721.1"
/db_xref="GI:3661491"
/translation="MPLLDKLEQYGVGVPCSELHAPSTYYHCOQRHHPDKRSARA
QRDWLKXIQRYDENHQYGVKVRQRLREGIRVACTVARLMVANGIAGVRGK
KVRTSAGPLQVTA"
/gene="r13"
/complement(3014..3340)
/gene="r13"
/complement(3014..3340)
/note="12.7-kDa protein"
/codon_start=1
/transl_table=11
/product="R13"
/protein_id="AAC61722.1"
/db_xref="GI:3661492"
/translation="MTKNTRFSPVRORAVRMVLESODEYDSORAAIGSIAPKTGCTP
ETLRVAIRQHERDTGGDGLTAERQRLKELERENRELRSNDILRQASAYFAAEF
DLRWKK"
/gene="r12"
/complement(4180..4656)
/gene="r12"
/pseudo
/gene="r12"
/psuedo
/gene="r12"
/note="transposase R12; ISAE1"
/pseudo
/codon_start=1
/transl_table=11
/5058..5372
/gene="r11"
/5058..5372
/gene="r11"
/note="possible precursor polypeptide"
/codon_start=1
/transl_table=11
/product="R11"
/protein_id="AAC61723.1"
/db_xref="GI:3661493"
/translation="MKHTGNEAYCCSVLPRKXQGVATYPPCSQGNVNSRLFL
STPSTQARSYNFTSVIQNTKIQNLIIQIIFLILNKSILQAYSEGSMCSIANIKRN
CS"
/gene="r10"
/5369..5629
/gene="r10"
/5369..5629
/gene="r10"
/note="hypothetical 8.6-kDa protein in dinGrarB 3' region"
/codon_start=1
/transl_table=11
/product="R10"
/protein_id="AAC61724.1"
/db_xref="GI:3661494"
/translation="MNKVICTTIVLGLTLTGATAADEINQDVVGLMPSDYISVGN
LTPPEVATALKDAEQAGSGHSRITSLSGKKKLDGTALIR"
/gene="f200"
/6272..6700
/gene="f200"
/6272..6700
/gene="f200"
/codon_start=1
/transl_table=11
/product="F200"
/protein_id="AAC61725.1"
/db_xref="GI:3661495"
/translation="MKKFINVYIKSDYDIIIVLRSLVLIIFMLFGYKWFDFEVMALKP
LISPTWLNIIYMFVGHGSGYFLGIVETITFLALISGFRPAAGILGDIIVITGLVT
LSLLPOLGVDSVFIKVDLLGAGSVLLKVDLKNHLNKN"
/complement(6984..7256)
/gene="r8"
/complement(6984..7256)
/gene="r8"
/codon_start=1

```

```

/transl_table=11
/product="neurotensin receptor R8"
/protein_id="AAC61726.1"
/db_xref="GI:3661496"
/translation="MTDITDRHTLRMSWSELFTAAQEAERFQDYERARIVMSFALHV
ATTINKNLAIAHIRCDTLLHKSKTVPGNNTGSRVCLRPQHPRR"
/gene="r7"
/7771..8019
/gene="r7"
/7771..8019
/gene="r7"
/codon_start=1
/transl_table=11
/product="maltopentaose forming amylase R7"
/protein_id="AAC61727.1"
/db_xref="GI:3661497"
/translation="MCRMSRWPNAPSRLLIGWELCCYLSWCHSHFNELNVAQQAQSAQT
VVAATAASCTFAVALLPPDTPCRWRFEADTSGFIFTG"
/gene="r6"
/8771..10237
/gene="r6"
/8771..10237
/note="similar to transposase of Chelatobactor heintzei"
/codon_start=1
/transl_table=11
/product="transposase R6"
/protein_id="AAC61728.1"
/db_xref="GI:3661498"
/translation="MTQKKGITQIHISAMKAGISVRSGRRIEKGWAKNSVRHWRTRK
DPLEAVWDSVLVLLKERPALTPPTLLEMLQDKVPGQYDPSLRRTWQRVREWKLOYG
AEQVWPRQRHQGLGLSDFTLKGWVVTIAGKLLAHKLYHPELEMSHWNRVYLG
GESFSALEGLQALGGLGVPVEHKTDLSLAARKQGGDGRRELTERRAALCOHQGM
QGVHNAAGRHENGVSVAHGLKRIICQALILRSGNDSTIEEYQAFITQQVRRNR
NQDLVKEERLHLKPLRSDYDELTVRSRSTINVKVSYSPSLVGLQLLRVR
LWDDRLSCYVSGSEVMSCPRVPEKGTARRIDFRHVIDSLAKKPGAFCHALRNDI
LPDDEWRLWRRLCNHLEPDMAGLMAVHALKLAAGYDDISVAKMEQMLNTPGNVDL
HLMRFLGIKEKALPVNVKQNLSSVEQLLRKGGSQ"
/gene="orfB"
/10399..10986
/gene="orfB"
/10399..10986
/note="similar to hypothetical protein B of Bacillus"
Query Match 97.98; Score 2046.8; DB 1; Length 13710;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 2051; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 34 CCCTGTCTCGGATTTTCAGCCAGCAGCATAGCTGCTGCAGAGGATGTGATGTTCTCG 93
DB 13710 CCCTGTCTCGGATTTTCAGCCAGCAGCATAGCTGCTGCAGAGGATGTGATGTTCTCG 13651
QY 94 GCATCCGGCTATGAGAAAAAGCTGACTAACCGACGCCCGCAGTCTTCTGTGTATGCCAG 153
DB 13650 GCATCCGGCTATGAGAAAAAGCTGACTAACCGACGCCCGCAGTCTTCTGTGTATGCCAG 13591
QY 154 GAGGAATTGAGTCCACCCAGTACCCAGCATCTCGCGGAGGCTCTGAGATCAGTAGAGGGT 213
DB 13590 GAGGAATTGAGTCCACCCAGTACCCAGCATCTCGCGGAGGCTCTGAGATCAGTAGAGGGT 13531
QY 214 GTGATGTTGAAAGTGTGTACGGGTAAACCGGAGGGCTGGAATATCAGCATCCAGGAATG 273
DB 13530 GTGATGTTGAAAGTGTGTACGGGTAAACCGGAGGGCTGGAATATCAGCATCCAGGAATG 13471
QY 274 CCAGCAGTTACACGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 333
DB 13470 CCAGCAGTTACACGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 13411
QY 334 ACTCCCAACGGTTTTTCTGCCATGAATACCGGGTTATGCCCTCTGCGCCGCCATTGAG 393
DB 13410 ACTCCCAACGGTTTTTCTGCCATGAATACCGGGTTATGCCCTCTGCGCCGCCATTGAG 13351
QY 394 CGTATTGAGGTTATCAGGGGGCGGATGTCACACTGTATGGCTCTGATGGATGGGCGGT 453
DB 13350 CGTATTGAGGTTATCAGGGGGCGGATGTCACACTGTATGGCTCTGATGGATGGGCGGT 13291

```

QY 454 GTGGTGAATATCATACAGAAAGAAATGCAGCAAAATGGCTCTCTTCGGTCAATGCAGGG 513
 DB 13290 GTGGTGAATATCATACAGAAAGAAATGCAGCAAAATGGCTCTCTTCGGTCAATGCAGGG 13231
 QY 514 CTGAATCTGCAGAAAGCAAAATGGGGTAAACAGCAGCCAGTTTAAATTTCTGGAGCAGT 573
 DB 13230 CTGAATCTGCAGAAAGCAAAATGGGGTAAACAGCAGCCAGTTTAAATTTCTGGAGCAGT 13171
 QY 574 GGTCCCTTGTGGATGATCTGTGAGCTGCGAGTACGGGTAGCAGCAACAGGTCAG 633
 DB 13170 GGTCCCTTGTGGATGATCTGTGAGCTGCGAGTACGGGTAGCAGCAACAGGTCAG 13111
 QY 634 GGTTCATCGGTACATCACTGAGCGATACAGCAGCAGCGTATTCCTTATCCACCGAG 693
 DB 13110 GGTTCATCGGTACATCACTGAGCGATACAGCAGCAGCGTATTCCTTATCCACCGAG 13051
 QY 694 TCACAGAAATTAATCTTGGTGCACGCTTGTGAGTGGAGGGTTCGAGCAGAGATGTGCTC 753
 DB 13050 TCACAGAAATTAATCTTGGTGCACGCTTGTGAGTGGAGGGTTCGAGCAGAGATGTGCTC 12991
 QY 754 TGGTTTGATATGATATGATACCCCGGAGCGTATGATTAACCGGGATGGGCACTGGGGAGT 813
 DB 12990 TGGTTTGATATGATATGATACCCCGGAGCGTATGATTAACCGGGATGGGCACTGGGGAGT 12931
 QY 814 CTGACGGGGGATATGACCGGACCTCGCTATGACGCAAAATTCAGCTGGCTAT 873
 DB 12930 CTGACGGGGGATATGACCGGACCTCGCTATGACGCAAAATTCAGCTGGCTAT 12871
 QY 874 GATCATCTTTCACCTTCGGAATGGAATCGTATCTGAACTGGAACGAGCAGAGAAAT 933
 DB 12870 GATCATCTTTCACCTTCGGAATGGAATCGTATCTGAACTGGAACGAGCAGAGAAAT 12811
 QY 934 AAAGTCTGAGCTTGTACGAGTGTACTGAAGCGGCAAAATGGGGCTTTCGGGTGAG 993
 DB 12810 AAAGTCTGAGCTTGTACGAGTGTACTGAAGCGGCAAAATGGGGCTTTCGGGTGAG 12751
 QY 994 CCGGGAGCTTAAGAAATCGAACTTATCTGAAATTCATTAAGCTTACCGCTTGGGA 1053
 DB 12750 CCGGGAGCTTAAGAAATCGAACTTATCTGAAATTCATTAAGCTTACCGCTTGGGA 12691
 QY 1054 GAATCTCATCTGTTACGGTGGGGGAGTTTCAGAGCTCTCCATGAAGACGAGTT 1113
 DB 12690 GAATCTCATCTGTTACGGTGGGGGAGTTTCAGAGCTCTCCATGAAGACGAGTT 12631
 QY 1114 GTCTTTCGACGACAGGTGAATCTTTCGGGCAAAAGCTGCTGGTATTTGCTGAGGAT 1173
 DB 12630 GTCTTTCGACGACAGGTGAATCTTTCGGGCAAAAGCTGCTGGTATTTGCTGAGGAT 12571
 QY 1174 GAGTGCATCTCAGGATCACTTGGCTGACTGGGCGAGCGCTATGAACATCATGAG 1233
 DB 12570 GAGTGCATCTCAGGATCACTTGGCTGACTGGGCGAGCGCTATGAACATCATGAG 12511
 QY 1234 CAATTCGGGGGACATTCAGTCCGGTGCATATCTGGTCTGGGATGTGGCAGATCCCTGG 1293
 DB 12510 CAATTCGGGGGACATTCAGTCCGGTGCATATCTGGTCTGGGATGTGGCAGATCCCTGG 12451
 QY 1294 AGCTGAAGGGGCTGTGACACGGGATATAGGCAACAGAAATGGGGCAGCTACATAAA 1353
 DB 12450 AGCTGAAGGGGCTGTGACACGGGATATAGGCAACAGAAATGGGGCAGCTACATAAA 12391
 QY 1354 GGGATTAGTGTGTGTCGGGCGAGGAAACAAATCTACTTGGTAAACCCGACCTGAAG 1413
 DB 12390 GGGATTAGTGTGTGTCGGGCGAGGAAACAAATCTACTTGGTAAACCCGACCTGAAG 12331
 QY 1414 CCGGAAGAGAGCGCTCAGTTATAGGCTGGGGTGTATTACGATPACCCCGCGGTCTGAAT 1473
 DB 12330 CCGGAAGAGAGCGCTCAGTTATAGGCTGGGGTGTATTACGATPACCCCGCGGTCTGAAT 12271
 QY 1474 GCAATGTACAGGTTTATGACTGACTTCCACAGATGTCTCTATTCCATAAT 1533
 DB 12270 GCAATGTACAGGTTTATGACTGACTTCCACAGATGTCTCTATTCCATAAT 12211

QY 1534 GATAACACCAATAGCTATGTAAACAGCGGAAGCCCGGTTGCACGGTGTGGAAATTGGC 1593
 DB 12210 GATAACACCAATAGCTATGTAAACAGCGGAAGCCCGGTTGCATGGTGTGGAAATTGGC 12151
 QY 1594 GGCACATTCGCCGTGTGGTTCAGAGGATGTACGCTGTCTACTGATTTACACCTGGAGCCGA 1653
 DB 12150 GGCACATTCGCCGTGTGGTTCAGAGGATGTACGCTGTCTACTGATTTACACCTGGAGCCGA 12091
 QY 1654 AGTGAACCAACGTGATGTGTATGATAACAAAGGTGCCCGCTGAGTTATACCCCTGAACACATG 1713
 DB 12090 AGTGAACCAACGTGATGTGTATGATAACAAAGGTGCCCGCTGAGTTATACCCCTGAACACATG 12031
 QY 1714 GTGAATGCGAAATCTGAATCTGGCAGATCAACGAAAGGTGGCAGTATGCTGGGTGCCCGT 1773
 DB 12030 GTGAATGCGAAATCTGAATCTGGCAGATCAACGAAAGGTGGCAGTATGCTGGGTGCCCGT 11971
 QY 1774 TATCGGGGAAACACACCGTTTACCCAGAAATTTTCGCTACTGAGCGCTGTACAGAAAG 1833
 DB 11970 TATCGGGGAAACACACCGTTTACCCAGAAATTTTCGCTACTGAGCGCTGTACAGAAAG 11911
 QY 1834 AAAGTGTATGATGAGAAAGAGAAATACCTGAAAGCTGGACGCTGGTGTGATGACGCTCTG 1893
 DB 11910 AAAGTGTATGATGAGAAAGAGAAATACCTGAAAGCTGGACGCTGGTGTGATGACGCTCTG 11851
 QY 1894 TCGTGGAGATGACGGATGCCCTGACGCTGAATGCTGCGGTGAATACCTGCTCAACAAG 1953
 DB 11850 TCGTGGAGATGACGGATGCCCTGACGCTGAATGCTGCGGTGAATACCTGCTCAACAAG 11791
 QY 1954 GATTACAGTACGCTGAGCCCTGTACAGTCCCGGTGAAGAGTACGCTGTATGCGGTGATTAC 2013
 DB 11790 GATTACAGTACGCTGAGCCCTGTACAGTCCCGGTGAAGAGTACGCTGTATGCGGTGATTAC 11731
 QY 2014 TTCAGACGGGATCATCAACAGAGTATGTATACCTAGCAGGAAATTAATGATGTCG 2073
 DB 11730 TTCAGACGGGATCATCAACAGAGTATGTATACCTAGCAGGAAATTAATGATGTCG 11671
 QY 2074 CTGAACATCATGTTCTGA 2091
 DB 11670 CTGAACATCATGTTCTGA 11653

RESULT 12
 LOCUS ECO278144
 DEFINITION Escherichia coli partial genomic DNA for pathogenicity island,
 shiga toxin producing strain 4797/97.
 ACCESSION AJ278144
 VERSION AJ278144.1 GI:14252925
 KEYWORDS int gene; integrase; ORF10; ORF11; ORF12; ORF13; orf15; ORF16;
 ORF17; ORF18; ORF19; ORF4; ORF5; ORF6; ORF7; ORF8; ORF9; putative
 sodium galactoside symporter; yicJ gene.
 SOURCE Escherichia coli
 ORGANISM Escherichia coli
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Escherichia.
 REFERENCE 1
 AUTHORS Schmidt, H., Zhang, W.L., Hemmrich, U., Jelacic, S., Brunder, W.,
 Tarr, P.I., Dobrindt, U., Hacker, J. and Karch, H.
 TITLE Identification and characterization of a novel genomic island
 integrated at selC in locus of enterocyte effacement-negative,
 Shiga toxin-producing Escherichia coli
 Infect. Immun. 69 (11), 6863-6873 (2001)
 JOURNAL
 MEDLINE 21481963
 PUBMED 11598060
 REFERENCE 2 (bases 1 to 37710)
 AUTHORS Zhang, W.L.
 TITLE Direct Submission
 JOURNAL Submitted (22-MAY-2000) Zhang W.L., Institut fuer Hygiene und
 Mikrobiologie, Universitaet Wuerzburg, Josef-Schneider-Str. 2,
 D-97080, GERMANY
 FEATURES
 source 1..37710
 /organism="Escherichia coli"


```
gene
/mol_type="genomic DNA"
/strain="4797/97"
/variety="serotype 103:H-"
/db_xref="taxon:562"
/country="Germany"
/notes="STEC from an asymptomatic carrier"
complement(1215..2654)
/genes="yicJ"
complement(1215..2654)
/genes="yicJ"
/codon_start=1
/transl_table=11
/product="putative sodium galactoside symporter"
/protein_id="CAC39281.1"
/db_xref="GI:14252926"
/db_xref="GOA:Q93588"
/db_xref="SPTREMBL:Q93588"
/translations="MVYKRYRSHQHPYSIKEQMKSEVLSSVKEKIGYGMGDAASHIIF
DNVLMYMPFYTDIFGIPAGFVGTMPFLVARALDAISDPQWGLLADRTSRMGKFRPW
LFCALPGVIVCLAYSTPDLXNGKMYALTYTLTLTYVNIYPYCALGCVITNDP
TQRISQSRFVLATAGMLSTVLMPLVNLIGGDKPLGROGGIAYLVSWAPMLAF
CFTTKERVEAPPTTSMREDLURDQWQDQWRIVGLLTFNILLVCVRGGAQMYVTV
ILGTPVFAFLTYCVGLNISALAKPLTDWKCKVTIIFWNTNALLAVISLAFVPM
QASITMVFVIFVIGVHLQVPTIQWVMSDVTVDYGEWCNKRLTGISFAGTFLVILG
LAFGGALIGMLAYGGYDAEAKAQNATISIIITFTIVPAICYLLSAIIAKRYYSLT
TNLKTVMEOQLAQKRRCCQOQFTSQEVQN"
2889..2983
/genes="selC"
2889..2983
/genes="selC"
/notes="trnA"
3284..4468
/genes="int"
3284..4468
/genes="int"
/codon_start=1
/transl_table=11
/product="integrase"
/protein_id="CAC39282.1"
/db_xref="GI:14252927"
/db_xref="GOA:Q93587"
/db_xref="SPTREMBL:Q93587"
/translations="MALTDIKVTKAPKDKPKVLADGGQWYLLINTGSKYVMKYRF
AGKEMLSIGVYDPDVTADAREKRSARKLLAAGDGPGEAKKEKIAQOMSLKNTFEA
IAEMQKXADRLRYRDEIIDTEKDIFFYIGKRPYAEIKPMELLALRMEKRGGA
LEKRVKVRQCEVFRVAVITGRADYNPAPDLASALATPKVHFPLFLANLPHPLND
LAGYTSIIKTATQIIMLTGVRTQELRFAWEDIDFETKLEIIPAEVYMKRPHIVP
LSQVIMLPKQLEPISKHPLVFIGENDPRKPIKESINOVIELIGYKRLTGHCGRH
TWSITLHEQGFNSATEMQLAHVDKMSIRGTYNHLYLDGREMKQWADYIDLSLSR
ES"
complement(4802..5683)
/notes="ORF4"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAC39283.1"
/db_xref="GI:14252928"
/db_xref="GOA:Q93586"
/db_xref="SPTREMBL:Q93586"
/translations="MDSACALVANGOGITLVSRILGSRAQLAQRMTSGNKPNRRR
RCEADAADAVLARDIISDMPTGYRRVWAILRQSRNKGVPFVNAKRVYRIMSENN
LILLKPKRPORREHGRISUKESQRCWSDGFEFGDDGAKLRVTFALDCCDEAID
WAASTGYKATQVQVMLGALIEKRGDKGAERPIQWLNDNGSAYTAHETROPARELNL
EPCITVASPQSNMAERFVTKMEDYITAFNPKPNVRTALHNLAVAHIEHYETHPHSA
LGYSLEPRYRQRTVST"
6567..7358
/notes="ORF5"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAC39284.1"
/db_xref="GI:14252929"
/db_xref="GOA:Q93585"
CDS
/mol_type="genomic DNA"
/strain="4797/97"
/variety="serotype 103:H-"
/db_xref="taxon:562"
/country="Germany"
/notes="STEC from an asymptomatic carrier"
complement(1215..2654)
/genes="yicJ"
complement(1215..2654)
/genes="yicJ"
/codon_start=1
/transl_table=11
/product="putative sodium galactoside symporter"
/protein_id="CAC39281.1"
/db_xref="GI:14252926"
/db_xref="GOA:Q93588"
/db_xref="SPTREMBL:Q93588"
/translations="MVYKRYRSHQHPYSIKEQMKSEVLSSVKEKIGYGMGDAASHIIF
DNVLMYMPFYTDIFGIPAGFVGTMPFLVARALDAISDPQWGLLADRTSRMGKFRPW
LFCALPGVIVCLAYSTPDLXNGKMYALTYTLTLTYVNIYPYCALGCVITNDP
TQRISQSRFVLATAGMLSTVLMPLVNLIGGDKPLGROGGIAYLVSWAPMLAF
CFTTKERVEAPPTTSMREDLURDQWQDQWRIVGLLTFNILLVCVRGGAQMYVTV
ILGTPVFAFLTYCVGLNISALAKPLTDWKCKVTIIFWNTNALLAVISLAFVPM
QASITMVFVIFVIGVHLQVPTIQWVMSDVTVDYGEWCNKRLTGISFAGTFLVILG
LAFGGALIGMLAYGGYDAEAKAQNATISIIITFTIVPAICYLLSAIIAKRYYSLT
TNLKTVMEOQLAQKRRCCQOQFTSQEVQN"
2889..2983
/genes="selC"
2889..2983
/genes="selC"
/notes="trnA"
3284..4468
/genes="int"
3284..4468
/genes="int"
/codon_start=1
/transl_table=11
/product="integrase"
/protein_id="CAC39282.1"
/db_xref="GI:14252927"
/db_xref="GOA:Q93587"
/db_xref="SPTREMBL:Q93587"
/translations="MALTDIKVTKAPKDKPKVLADGGQWYLLINTGSKYVMKYRF
AGKEMLSIGVYDPDVTADAREKRSARKLLAAGDGPGEAKKEKIAQOMSLKNTFEA
IAEMQKXADRLRYRDEIIDTEKDIFFYIGKRPYAEIKPMELLALRMEKRGGA
LEKRVKVRQCEVFRVAVITGRADYNPAPDLASALATPKVHFPLFLANLPHPLND
LAGYTSIIKTATQIIMLTGVRTQELRFAWEDIDFETKLEIIPAEVYMKRPHIVP
LSQVIMLPKQLEPISKHPLVFIGENDPRKPIKESINOVIELIGYKRLTGHCGRH
TWSITLHEQGFNSATEMQLAHVDKMSIRGTYNHLYLDGREMKQWADYIDLSLSR
ES"
complement(4802..5683)
/notes="ORF4"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAC39283.1"
/db_xref="GI:14252928"
/db_xref="GOA:Q93586"
/db_xref="SPTREMBL:Q93586"
/translations="MDSACALVANGOGITLVSRILGSRAQLAQRMTSGNKPNRRR
RCEADAADAVLARDIISDMPTGYRRVWAILRQSRNKGVPFVNAKRVYRIMSENN
LILLKPKRPORREHGRISUKESQRCWSDGFEFGDDGAKLRVTFALDCCDEAID
WAASTGYKATQVQVMLGALIEKRGDKGAERPIQWLNDNGSAYTAHETROPARELNL
EPCITVASPQSNMAERFVTKMEDYITAFNPKPNVRTALHNLAVAHIEHYETHPHSA
LGYSLEPRYRQRTVST"
6567..7358
/notes="ORF5"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAC39284.1"
/db_xref="GI:14252929"
/db_xref="GOA:Q93585"
CDS
/db_xref="SPTREMBL:Q93585"
/translations="MDGRRSRHTDDTVLLRIHHVIGELPTYGYRRVWALLRRQAEJD
GMPAINAKRYRIMRNQALLERKPAVPPSKRAHTGRVAVKESNQWCSGDFBFRCDN
GEQLRVTFALDCCDREALHWAIVTGFNSETQVQVMLGAVERRFGLNLPASVPEWLTG
NGSCYRANETRQFARMGLPKSTAVRSPESNGIAESFVKTIKRDYIISVMPKPDGLTA
AKNLAEAFEHVNEWHPSALGYRSPREYLRQAQASNGLSNDRCLFI"
complement(8616..10154)
/notes="ORF6"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAC39285.1"
/db_xref="GI:14252930"
/db_xref="SPTREMBL:Q93584"
/translations="MNDISDDIFLLKQRLAEQALIHAEQKLSNRERIIDLQQL
DKLRMNFGRSRKSVSRRIAQMAGADLNLRQESDITLGRVYDPAVORPLRTRKRF
PESLPREDKRLPAAPCCPGGSLSGEDTAELQELMELMSAFVRVIRVREKHACTQC
DAIVQAPAPSRPDRGIAGPGLLARLYAEHTPLRQSEIYGRQGVVLRSLSSG
WDACCRLSPLEALHGVYMDKHLHADDTPQVLLPGNKTKTKRLWAYVRDRNA
GSALAPVAPAYSDRKGIIHPOTHLACFSGLVADAVAGFNELRYNGGITTEACMAHA
RRKHVDVHVPISALFEALEQIGOLYAIETADIEGMPAEQRLAERKPKPLKSLSES
WLRKMKILSRHSELAKAFAYALNQMPALTYANDGVEIDNINNAENALNALSGLRAN
FLPFGSDHGGERGALLYSLTGKLNDDVPESYLRHLVGLVADWPVNRVSELLPWRIA
LPAB"
complement(11248..15339)
/functions="protease"
/notes="ORF7"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAC39286.1"
/db_xref="GI:14252931"
/db_xref="GOA:Q93583"
/db_xref="SPTREMBL:Q93583"
/translations="MKKIYALKYSSLTGGLIYNSLKKVTGKTRRLMTVSLVLSVT
LSALPGKASTVSAEIPYQTPRDAENKGVPIGVTGIEINNNNGKVGVLDPMLDPS
SLRDGHHTLLHPGYVYVYAKHGLQSVSSATGYDQIYKIVDNNLAGIDFSAFLNKL
VTEVIPADIQKDKFNRRYAFYRAGVSGSYIRYANGTDKLLQAYTPEKAYLTGTV
LKNPYTHNGKMIISANPNTDKNQGLASVQSGSDGSPLYAWDNIKWKWLAGVT
LHNVYKVGARDWLLIPHDPIISQILODLPKPIIVASPEENILRWRDRSGTGLSQG
EKIPLMTGVSNGNANTGNLVFSENGKIELSVSEQAGYLOFDKDYTLVNTNSTW
TGAGIIVGDANVNGVNGIAGNLHKVSGSLTVANGHNGKGLDNGVYFTSRLOQA
DAKQKQAPSHINIASGATVKLNGANQVDADNISWYRGCKLKLNGVYFTSRLOQA
DYGAESINDQDKSIYLSPLKAEINVVNNINMGTKGKGLDLYITFDGNYT
LLKSNRYSGALFALANNQSEWRLKDKKEKALGLYTKMKVQESAPLSYHKGNTY
SVIPLKAGNDILTDGNSISGDMKQDGLIFQGPVTHAGCTVSAQSDWENREF
SLANLANNADFLSRNFWNGIRAVNOSTVIGDVTVDKNDGLDNDVSVEGKS
AAAGTSYTGHTLEOKSALDIRNFRGVTSEDSHINSSSVSLDASSFINSLSL
IHKGKALTAGGLFTSGSDIGDASLLLTGTPVNSDDAARLPTINNADGFKLMSDS
VLKAROQSVVGDIIISQKQATISFETESGREGMLSEKASRGLAVGLSFTAYRGA
HAPSASATMNTWLTGDSLSRLSKNTGSMYFTTGAANKAFHTLTVDELITNGTAY
AMRTDLKADLVNQLSKDNILLVDFLNKPTGKLDIEIVASPGNSKDVFKGSE
QAIGFSNVPTVITAKDAGRTTNLTGYRMAENPAATQASGLASVGYKSFSEVNNL
NKRMDLRDINGEAGAWARIMSGTSAGAGFSDNTHVQVGVKKEHLDGLDFTGFT
VTHDSASADAEGKTKSVAGLYASAMFDSGAVIDLIGKVYVHNDNEYTATFAGGI
RDYSTHWSYAGAGRGYRCHVTEDTIEWPQALYVGAISGKQFAWKQGHLSMKRDY
NPLIGRTGVDSGKSGKMDKPTARAGLGTFQFOLLANGETVLKADSGEKRIKGEKUSR
MLMSVGLNAEIRDNVFRGLEFEKSAFGKYNVDVANFRYSF"
complement(16252..16989)
/functions="unknown"
/notes="ORF8"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAC39287.1"
/db_xref="GI:14252932"
/db_xref="GOA:Q93582"
/db_xref="SPTREMBL:Q93582"
/translations="MIVCVVILNRPPEIAGKCLFQKNDVLLTDNTSQSFFLRHP
GNVTELRIDECIVRYMESADRIQHSDSAVSVPPIKTFSTHIALHNVWQNLISKTL
POQFSQMFACLSIFSTHGRNFFLFCINSVALKRVYIICSSPSRAWRVDISNLS
YMSESLKRLKQBEKYSFQIILLDIRMHNARQLISSKLSVNQIAIRCGYASTPYFISV
FREYFGITTHQMINQSQGFIESRIAK"
```


CDS

```

17821..19908
/function="adhesion"
/feature="ORF9"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAC39288.1"
/db_xref="GI:14252933"
/db_xref="GOA:O93SE1"
/db_xref="SPTREMBL:O93SE1"
/translation="NR1TILASVIPCIFPSASSMAEDVMIVSASGVEKLTNAAS
VSVISEELQSQXHDLAELRAVDEGVESGTHGGLLEISIRGMPASYLILIDGV
RQGGSDVTPNGFSAMNTGFMPPLAAIERIEVIFPNSTLYSGDAMGVNIIITKKA
DKWLSVNAVNLQESNKGNSQFNFSGLPVDSDLSLOIRGSTQORGSVTSLS
DTSATRIPYTESQNYNLGARDKATQBDVLMFMDTRQYNDQDQLGSLGTGYD
HTLVRNKISAGYDHTFTFTGWKSYLNWNETENKRELVSLSKDKWGLAGPREL
KZNLILNSLTLPLGESHLLTVGEYQSSNMKGQVLAISTGETPROKSWVFAEDW
HUTDIALTAGRLRYEHHQFGHFSFRAYLVWDVTDANTLKGVTGTYKATMGQLKH
GISVGGQRTNLLGNPKPEESVSYEAGVYDNLGNLNIITGTFIDFNKLVSYIS
IDNTNSYTNSSGKARLHGVFAGTLPMSDEVTLSLNTYTQSEORDDNGKALSYT
PEHWNALNMQITBEVASWLGARYGKTPFRTRYSSLSAVQKRYDEKGEYLKAWT
VVDAGLSWKVTDALTNAAVNNLNKDYSDVLSYADSKSTLYAGDYFQTSSTGYVI
PERNWMISLNTQF"
complement(20254..22254)

```

CDS

```

Query Match      89.1%; Score 1863.8; DB 1; Length 37710;
Best Local Similarity 93.5%; Pred. No. 0;
Matches 1956; Conservative 0; Mismatches 132; Indels 3; Gaps 1;

QY      1  ATCGGAATACCACTCGGCTCCGTCAGTCCCTGCTCGGATTTTCAGCCAGCAGC 60
DB      17821 ATCGGAATACCACTCGGCTCCGTCAGTCCCTGCTCGGATTTTCAGCCAGCAGC 17880

QY      61  ATAGCTGCTGCAGAGGATGTGATGTTGTCGGCATCCGGCTATGAGAAAAGCTGACT 120
DB      17881 AT---GGCAGCAGAGGATGTGATGTTGTCGGCATCCGGCTATGAGAAAAGCTGACT 17937

QY      121 AACGACGCCAGCTGTTTCTGTGATTAGCCAGGAGGANTTCAGTCCAGCCAGTACCAC 180
DB      17938 AACGACGCCAGCTGTTTCTGTGATTAGCCAGGAGGANTTCAGTCCAGCCAGTACCAC 17997

QY      181 GATCTGGCGAGGCTCTGAGATCAGTAGAGGCTGTGATGTTGAAAGTGGTACGGGTAAA 240
DB      17998 GATCTGGCGAGGCACTGAGAGCTGTGGAGGCTGTGATGTTGAAAGTGGTACGGGTAAA 18057

QY      241 ACCGAGGCTGGAATACGATCCGAGGAATCCAGCCAGTTCACCGTGTATCTGATT 300
DB      18058 ACCGAGGCTGGAATACGATCCGAGGAATCCAGCCAGTTCACCGTGTATCTGATT 18117

QY      301 GATGGTGTTCGTCAGCGCGGAGCAGTGCAGTGCATCCCAACGGTTTTTCTGCCATGAT 360
DB      18118 GATGGTGTTCGTCAGCGCGGAGCAGTGCAGTGCATCCCAACGGTTTTTCTGCCATGAT 18177

QY      361 ACCGGGTTCATGCCCTCTGGCGGCAATTGAGCGTATTGAGTTATCAGGGGGCGGATG 420
DB      18178 ACAGGGTTCATGCCCTCTGGCGGCAATTGAGCGTATTGAGTTATCAGGGGGCGGATG 18237

QY      421 TCACACTGTATGGCTCTGATGCGATGGCGGTGTGTTGAATATCATTTACAGAGAGAT 480
DB      18238 TCACACTGTATGGCTCTGATGCGATGGCGGTGTGTTGAATATCATTTACAGAGAGAT 18297

QY      481 GCAGACAAATGGCTCTCTCCGTCATGACGGGCTGAATCTGCAGGAAACCAAAATGG 540
DB      18298 GCAGACAAATGGCTCTCTCCGTCATGACGGGCTGAATCTGCAGGAGAGTAAATGG 18357

QY      541 GGTAAACAGCAGCAGTTTAAATTTCTGAGCAGTGGTCCCTTGTGATGATGTTCTGTAC 600
DB      18358 GGCAACAGCAGCAGTTTAAATTTCTGAGCAGTGGTCCCTTGTGATGATGTTCTGTAC 18417

QY      601 CTCAGGTACCGGTAGCACACACAGCGTCAGGTTTCATCGTTCACATCAGTACGCGAT 660
DB      18418 CTCAGATCCGCGCAGCACACACAGCGTCCAGGCTCATCGTTCACATCAGTACGCGAC 18477

```

```

QY      661 ACAGCAGGCGCAGCTATTCCTTATCCACGAGTCAAGAAATTAATCTTGTGTGACGT 720
DB      18478 ACATCAGCAGCAGCTATTCCTTATCCACGAGTCAAGAAATTAATCTTGTGTGACGT 18537

QY      721 CTTGACTGGAAGCGTTCGAGCAGGAGTGTCTGTTTGTATGATGATACCAACCCGAC 780
DB      18538 ATTGACTGGAAGCGTTCGAGCAGGAGTGTCTGTTTGTATGATGATACCAACCCGAC 18597

QY      781 CGTTATGATAACCGGATGGCAACTGGGAGTCTGACGGGGGATATGACCGGACCGCTG 840
DB      18598 CGTTATGATAACCGGATGGCAACTGGGAGTCTGACGGGGGATATGACCACTACTCTG 18657

QY      841 CGCTATGAGCAACAAAATTTTTCAGCTGGCTATGATCATACTTTTACCTTCGGAACATGG 900
DB      18658 CGCTATGAGCAACAAAATTTTTCAGCTGGCTATGATCATACTTTTACCTTCGGAACATGG 18717

QY      901 AAATCGTATCTGAATCGAACGAGACAGAAAATAAAGTCTGAGCTTTGACGAGTGA 960
DB      18718 AAATCGTATCTGAATCGAACGAGACAGAAAATAAAGTCTGAGCTTTGACGAGTGA 18777

QY      961 CTGAAGCGCAGCAAAATGGGGCTTCCCGTTCAGCGCGGAGGACTTAAGGAATGAAACCTT 1020
DB      18778 CTGAAGCGCAGCAAAATGGGGCTTCCCGTTCAGCGCGGAGGACTTAAGGAATGAAACCTT 18837

QY      1021 ATCCTGAATTCATTACTGCTTACCCCTCTCGGAGAAATCTCATCTGTTTACGGTGGGGGC 1080
DB      18838 ATCCTGAATTCATTACTGCTTACCCCTCTCGGAGAAATCTCATCTGTTTACGGTGGGGGC 18897

QY      1081 GAGTTTCAGAGCTCGTCCATGAAGACGAGTGTGCTTCCCGACACAGGTGAACCTTTC 1140
DB      18898 GAGTTTCAGAGCTCGTCCATGAAGACGAGTGTGCTTCCCGACACAGGTGAACCTTTC 18957

QY      1141 CGGACAGAAAAGCTGGTGGTATTTGCTGAGGATGAGTGGCATCTCAGGGATGCACTTTGCG 1200
DB      18958 CGGACAGAAAAGCTGGTGGTATTTGCTGAGGATGAGTGGCATCTCAGGGATGCACTTTGCG 19017

QY      1201 CTGACTCGGGCAGCCCTATGAACATCATGAGCAATTCGGGGGACACTTCAGTCCCGCT 1260
DB      19018 CTGACTCGGGCAGCTCGTATGAACATCATGATCAATTCGGGGGACACTTCAGTCCCGCT 19077

QY      1261 GCATATCTGCTGCGGATGTGGCAGATGCCTGACGCTGAAGCGGTGAGCCAGCAGGGA 1320
DB      19078 GCATATCTGCTGCGGATGTGGCAGATGCCTGACGCTGAAGCGGTGAGCCAGCAGGGA 19137

QY      1321 TATAAGCACCACCAATGGGGCAGCTACATAAAGGATTAGTGTGTCTCGGCGACAGGGA 1380
DB      19138 TATAAGCACCACCAATGGGGCAGCTACATAAAGGATTAGTGTGTCTCGGCGACAGGGA 19197

QY      1381 AAAACAAATCTACTTGTGTAACCCCGACTGACCCGAACTGAAGCGGAGAGAGTGAAGTTATGAGCGT 1440
DB      19198 CGTACTAATCTACTTGTGTAACCCCGAACTGAAGCGGAGAGAGTGAAGTTATGAGCGT 19257

QY      1441 GGGGTGTATTACGATAACCCCGGCTGTAATGCCAATGTACAGGTTTTTATCAGCTGAC 1500
DB      19258 GGGGTGTATTACGATAACCCCGGCTGTAATGCCAATGTACAGGTTTTTATCAGCTGAC 19317

QY      1501 TTCTCCAAACAGATGTTCTTATTTCATTAATATTAACCAATAGCTATGTAACAGC 1560
DB      19318 TTCTCCAAACAGATGTTCTTATTTCATTAATATTAACCAATAGCTATGTAACAGC 19377

QY      1561 GGAAGGCCCGGTTGACGGTGTGGAATTTCCCGGCACTTCCCGCTGTGTGTGAGAGAT 1620
DB      19378 GGAAGGCCCGGTTGACGGTGTGGAATTTCCCGGCACTTCCCGCTGTGTGTGAGAGAT 19437

QY      1621 GTCACGCTGTCACTGAATTTACACTGGACCGAAGTGAACAACTGATGTTGATGAACAAA 1680
DB      19438 GTCACGCTGTCACTGAATTTACACTGGACCGAAGTGAACAACTGATGTTGATGAACAAA 19497

QY      1681 GGTGCGCGCTGAGTTATACCCCTGAACACATGTTGTAATGCGAACTGAACCTGCGAGTTC 1740
DB      19498 GGTGCGCGCTGAGTTATACCCCTGAACACATGTTGTAATGCGAACTGAACCTGCGAGTTC 19557

QY      1741 ACCGAGAGGTGGCATCATGCTGGTCCCGTTTATCGCGGGAACACACACGCTTTCACC 1800

```

```

Db 19558 ACCGAAAGGTTGGCATCATGCTGGGTCGCCGTTATCGCGGAAAAACACACGTTTCACC 19617
Qy 1801 CAGAAATTATTCCTCCTCAGCGCTGTACAGAAAGAGTGTATGATGAGAAAGGAGATAC 1860
Db 19618 CGGAATTATTCCTCCTCAGCGCTGTACAGAAAGAGTGTATGATGAGAAAGGAGATAT 19677
Qy 1861 CTGAAGCCTGGACGGTGGTGGATGCAAGTCTGTCTGCGGAGAGTACGAGATGCCCTGACG 1920
Db 19678 CTGAAGCCTGGACGGTGGTGGATGCAAGTCTGTCTGCGGAGAGTACGAGATGCCCTGACG 19737
Qy 1921 CTGAATCTGCGGTTGAATACCTCTCAACAAGATTACAGTACAGTACGAGCTGTACAGT 1980
Db 19738 CTGAATCTGCGGTTGAATACCTCTCAACAAGATTACAGTACAGTACGAGCTGTACAGT 19797
Qy 1981 GCCGGTGAAGTAGCTGTATGCGGTTGATTAATTCAGACGGGATCATCAACAAGCA 2040
Db 19798 GCCGATAAGAGTAGCTGTATGCGGTTGATTAATTCAGACGGGATCATCAACAAGCA 19857
Qy 2041 TATGTGATCTGAGCGAATTAATTCAGATCTGCTGATCTGCTGATCTGCTGATCTGCTG 2091
Db 19858 TATGTGATCTGAGCGAATTAATTCAGATCTGCTGATCTGCTGATCTGCTGATCTGCTG 19908

RESULT 13
AF447814 71684 bp DNA linear BCT 16-OCT-2002
LOCUS Escherichia coli strain CFT073 pathogenicity island II, partial
DEFINITION
Sequence.
VERSION AF447814.1 GI:18265859
KEYWORDS
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE
1 (bases 1 to 71684)
Rasko, D.A., Phillips, J.A., Li, X. and Mobley, H.L.
Identification of DNA sequences from a second pathogenicity island
of uropathogenic Escherichia coli CFT073: probes specific for
uropathogenic populations
J. Infect. Dis. 184 (6), 1041-1049 (2001)
21458620
MEDLINE
PUBMED
11574920
REFERENCE
2 (bases 1 to 71684)
Rasko, D.A., Phillips, J.A., Li, X. and Mobley, H.L.T.
Direct Submission
Submitted (14-NOV-2001) Dept of Microbiology and Immunology,
University of Maryland School of Medicine, 655 W. Baltimore Street,
Baltimore, MD 21201, USA
FEATURES
Location/Qualifiers
1..71684
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="CFT073"
/db_xref="taxon:562"
/notes="pathogenicity island II; isolated from the blood of
a Homo sapiens female; urinary tract pathogen"
<1..339
/notes="ORF1; similar to Escherichia coli L0014 encoded by
GenBank Accession Number AF071034"
/codon_start=1
/transl_table=11
/product="L0014-like protein"
/protein_id="AAL67333.1"
/db_xref="GI:18265860"
/translation="GSHSHOVPSAGSWPALPICVNPSTWENRYNMCWRRPPOVTCI
SFVADGVTLILKYSGPLMYCASPDPNWRKASISGLRCVTVRYPLPARNWQSSISWTG
VSQKHSALTH"
339..1028
/notes="ORF2; similar to Escherichia coli L0015 encoded by
GenBank Accession Number AF071034"
/codon_start=1

```

```

/transl_table=11
/product="L0015-like protein"
/protein_id="AAL67334.1"
/db_xref="GI:18265861"
/translation="MPGRITKMSOKYLIRIAELESQKQDQOLSLVETKTFLLRS
ALAPAEKTEDEDEIEHLRAQIEKRLMFGTCEKLRREVQAEALLKQROSDPR
YSGREDQVPRQLRQSRHRPPFAHLPREIHLRLESESCPCGCGELDLGVSAEQ
LELVSSALKVIRTERVKKACTKDCICVEAPAPSRDRDRAWYRPRITTCPRVYGIKLRTP
ATVSSSE"
1102..1929
/notes="ORF3; similar to Escherichia coli L0015 encoded by
GenBank Accession Number AF071034"
/codon_start=1
/transl_table=11
/product="L0015-like protein"
/protein_id="AAL67335.1"
/db_xref="GI:18265862"
/translation="MTPLNDALYRYVNMTRKLHDDTPVKVLAFLGKTKTGRITVY
RDDNAGSSPPAVMPFVSPNROGQHPERGLQADAFGYDRLPSARREGGA
LVEYACCAHARRKTHDVYISKSMABEALKRISLXAIEDIRGLPESERLAVROOR
SKALLTSIHMEVKNKNTLSKSRLEAPFVYLNQDLCYRDDDLGLADNNTABRA
LRVLCGKNSYDLCQILSPRPYAPAGSLYKLNKQSDCILFTINSYSGILAP"
complement(2039..2782)
/notes="ORF4; similar to Agrobacterium tumefaciens
transposase encoded by GenBank Accession Number AF034769"
/codon_start=1
/transl_table=11
/product="putative transposase"
/protein_id="AAL67336.1"
/db_xref="GI:18265863"
/translation="MRDLIPARAARAKRDSRVQRILSMRLTRDQKHVAGHWTKGRT
WLANSPSQSOQIAFOHQCSQEQIEDRLQDQIRSLLEPWSLNCVQQLKQK
VGLIATLVAELGDRFSNPKQLMAFLGVPEYSSGNSIRPGITKVGSELRL
LYEAMSVYRTPAKVAGWLYVYRPNSTQVSKDIANKAQORLCSYRSLTAKGKQVA
ITAVARELTGFMDIALAASFSQKQN"
complement(2810..3214)
/notes="ORF5; similar to Agrobacterium tumefaciens
transposase encoded by GenBank Accession Number AF034769"
/codon_start=1
/transl_table=11
/product="putative transposase"
/protein_id="AAL67337.1"
/db_xref="GI:18265864"
/translation="MSENIAAYVGLDVHKETLVAIAAPERLGEVRYGTINNEAQA
VRRLFKQLGYNLISCEYAGPCGGLYHQLTAMNKKQVIAPSRIPKSPTRIND
HRDAISLARLLRAGELTPVMKGFQHTGGRY"
3611..4297
/notes="ORF6; similar to Shigella sonnei transposase
encoded by GenBank Accession Number AB029445"
/codon_start=1
/transl_table=11
/product="putative transposase"
/protein_id="AAL67338.1"
/db_xref="GI:18265865"
/translation="MRPIAPISDERBLMKAIHKTHDKNYABELTAMLMHLRGDRVS
DVARTLCARSSVGRINWFTQSGVGLKSLPACRARRWPFHEHCTILLRELVKHSQD
FGYQRSSVTELLAIKINEITCOLNAGTVPWPSAGIWRRAAPTLRLRDPKDEK
MAAIIKALDECCAEHPVFYDEVDIHLNPKIGADWQLRGQQRVTFPGQVKPAMWAAT
AKVRRCSAC"
complement(4678..5025)
/notes="ORF7; similar to Escherichia coli L0008 encoded by
GenBank Accession Number AF071034"
/codon_start=1
/transl_table=11
/product="L0008-like protein"
/protein_id="AAL67339.1"
/db_xref="GI:18265866"
/translation="MLPNSRQPTLHRLHLLPLTPYVRCSTAHITKDNAMLYSLSHOTN
DFGSESWIHTGTGYLLRLTDANSYPVLRKLRLGLSKTFRRLVVTLLRCYGVGLHLAA
SAGCLPLGPIFDW"
complement(4892..5116)
/notes="ORF8; similar to Escherichia coli L0008 encoded by
GenBank Accession Number AF071034"

```

CDS

CDS

CDS

CDS

```
/codon_start=1
/transl_table=11
/product="L0008-like protein"
/protein_id="AAL67340.1"
/db_xref="GI:18265867"
/translation="MGRIVVDVGDIDPAELNNKVAENGYSLSRVVDAPEQQTDSPPPF
TALTDIESLQRYTHGQRLAVLTVTSDE"
5289..6755
```

CDS

```
/note="ORF9; similar to Escherichia coli R6 protein
encoded by GenBank Accession Number AF081285"
```

```
/codon_start=1
/transl_table=11
/product="r6-like protein"
/protein_id="AAL67341.1"
/db_xref="GI:18265868"
/translation="MTQRKGIITQHISAMKAGISVRSGRRIKESWAKNSVRHWETRK
DPLEAVWDSMLVPLKRPALPTTLEMLQDKYPCQYNSLRRTWQRVRSWVRLQYV
AGEQVMPROHQPLGLSDFTVEHKVTVTIGKLAHLKLYHFLLEWSHWVRLVGL
GEFSALAGLOALGQGVPEVHKTDLSLRKAWKQGGDGRRELTERTVAALCOVGM
QGVHNAAGRHENGVSASAGHLKRRCQALLILGSDNSTIEEQAFITQVVMHNR
NQDLVKEERLHKPLPRRSADYDLELTWVRSSTINVKHVVSVFSLKVLGQLLRVR
LWDRUSCYVGSSEVMSCPKRPKERRIDFRHVIDSLAKKPGAPFCHATLRNDI
LPDEWRRLWRICNLHLPDMDAGRLMHALKLAAGYDDISVVAKGMSQMLNTPGNVDI
HRLMFLGIXEKALPVNVNKHNLSSYEQLLRGKGSQ"
complement(7582..7959)
```

CDS

```
/note="ORF10; similar to SwissProt Accession Number
P76365"
```

```
/codon_start=1
/transl_table=11
/product="intergenic-region protein"
/protein_id="AAL67342.1"
/db_xref="GI:18265869"
/translation="MKTLPDTHVREASRCPSPVTWOTLLTOLLDOHYGLTNDTPPA
DERVIOHIEAGSLCDVNFVLEKVALVTRTPQPSACTRSQQLNSIDILERRATG
LWTRSNRYMVNDIIRGESEAK"
complement(8006..8488)
```

CDS

```
/note="ORF11; similar to SwissProt Accession Number
P76364"
```

```
/codon_start=1
/transl_table=11
/product="intergenic-region protein"
/protein_id="AAL67343.1"
/db_xref="GI:18265870"
/translation="MRPFTSMHIPSXPVSTPADLSSLNYSHPICEDPIFVSDTLHET
NYDDNDRPWGLPCTVTPCFGARLVQEGNQLYLADRAGIRGOFSDADAYHLDQAF
FLMKQLEMLLTIGELNPRHQHTVLYARGLTCEADTLGSCGYVMVYPTLAPATTS
"
```

CDS

```
/note="ORF12; similar to SwissProt Accession Number
P76363"
```

```
/codon_start=1
/transl_table=11
/product="intergenic-region protein"
/protein_id="AAL67344.1"
/db_xref="GI:18265871"
/translation="MKIITRGEAMRIHQHQPASRLFPFCTGKYRWHGSTETVTGREVQ
DIPGLAVFAERRKDSFGPYVRLMSVTLN"
complement(9377..9869)
```

CDS

```
/note="ORF13; similar to SwissProt Accession Number
P76362"
```

```
/codon_start=1
/transl_table=11
/product="intergenic-region protein"
/protein_id="AAL67345.1"
/db_xref="GI:18265872"
/translation="WEACTMQQLSFLPGEMTTRSLIRALKTLDRHLHPGVAFPTS
TRAARWLLINWAGLEREFVRLNNQNLQIATGLTFTGINRTVEHPREVIKRALY
HNAAVVLAHNPSGEVTPSKADRLITERLQALGLVDIRVPDHLIVGGSQVFSFAEH
GLL"
9389..10348
```

CDS

```
/note="ORF15"
/codon_start=1
/transl_table=11
```

```
/product="unknown"
/protein_id="AAL67347.1"
/db_xref="GI:18265874"
/translation="MFGKGENLAATDYQMRHPDHIHQACLYQTFTRDKPVCLAGCDFT
GRVIMCOHGGGIVIOGAFNHFPMDFRAVDGAGEEGTGNQLIIVIOQHPELFTLQ
SRVIVQPPFRTGGGEGHAGFMKAVQGFQNEALSGGFGSQGKLKHCPLH
SVDDAHNGAFAFRMLQGVIPQTVIMNRHGIITLRTAGVTVRHQTGNGAGFRAHFAVT
```

```
Query Match          49.0%; Score 1025; DB 1; Length 71684;
Best Local Similarity 99.5%; Pred. No. 4,9e-297;
Matches 1028; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

```
QY 1 ATGCGAATAACCACTCTGGCTTCGGTAGTCATTCCTGTCTCGGATTTTCAGCCACGACG 60
Db 34034 ATGCGAATAACCACTCTGGCTTCGGTAGTCATTCCTGTCTCGGATTTTCAGCCACGACG 34093
```

```
QY 61 ATAGCTCTCCAGAGATGATGATGATGTCGGGATCCGGTATGAGAAAAGTGACT 120
Db 34094 ATAGCTCTCCAGAGATGATGATGATGTCGGGATCCGGTATGAGAAAAGTGACT 34153
```

```
QY 121 AACCGAGCCCGGAGTGTTCGTGTGATTATCCAGGAGGAATTCAGTCCAGCCAGTACCA 180
Db 34154 AACCGAGCCCGGAGTGTTCGTGTGATTATCCAGGAGGAATTCAGTCCAGCCAGTACCA 34213
```

```
QY 181 GATCTGCGGAGGCTCTGAGATCAGTAGAGGTGTGGATGTTGAAAGTGGTACGGGTAA 240
Db 34214 GATCTGCGGAGGCTCTGAGATCAGTAGAGGTGTGGATGTTGAAAGTGGTACGGGTAA 34273
```

```
QY 241 ACCGAGGGGTGGAATCAGCATCCAGGAATGCCAGCCAGTTACACGCTGATCTGATT 300
Db 34274 ACCGAGGGGTGGAATCAGCATCCAGGAATGCCAGCCAGTTACACGCTGATCTGATT 34333
```

```
QY 301 GATGGTCTTCGTCCAGCGGAGACGATGACGTGACTCCCAACGGTCTTCTGCCATGAT 360
Db 34334 GATGGTCTTCGTCCAGCGGAGACGATGACGTGACTCCCAACGGTCTTCTGCCATGAT 34393
```

```
QY 361 ACCGGGTTATGCCCTCTGCGCGCATTTAGCGTATTAGGTTATCAGGGGGCCGATG 420
Db 34394 ACCGGGTTATGCCCTCTGCGCGCATTTAGCGTATTAGGTTATCAGGGGGCCGATG 34453
```

```
QY 421 TCCACACTGTATGGCTCTCATGCGGTGGCGGTGGTGAATATCATACAGAAAGAT 480
Db 34454 TCCACACTGTATGGCTCTCATGCGGTGGCGGTGGTGAATATCATACAGAAAGAT 34513
```

```
QY 481 GCAGACAAATGGCTCTCTTCCTGTCATGCAATGCAAGGCTGAATCTCAGAGAAAGCAAAATGG 540
Db 34514 GCAGACAAATGGCTCTCTTCCTGTCATGCAATGCAAGGCTGAATCTCAGAGAAAGCAAAATGG 34573
```

```
QY 541 GSTAACAGAGCCAGCTTAAATTTCTGAGCAGTGGTCCCTTGTGGATGATTCTCTCAGC 600
Db 34574 GSTAACAGAGCCAGCTTAAATTTCTGAGCAGTGGTCCCTTGTGGATGATTCTCTCAGC 34633
```

```
QY 601 CTGCAAGTACGGGTAGCACACACAGCGTCAAGGTTTCATCGGTCAACATCATCTGAGCGAT 660
Db 34634 CTGCAAGTACGGGTAGCACACACAGCGTCAAGGTTTCATCGGTCAACATCATCTGAGCGAT 34693
```

```
QY 661 ACAGAGGACCGGTATTCCTTATCCACGAGTACAGAAATTAATCTTGGTGACATG 720
Db 34694 ACAGAGGACCGGTATTCCTTATCCACGAGTACAGAAATTAATCTTGGTGACATG 34753
```

```
QY 721 CTTGACTGGAAGCGCTCGAGCAGGATGTGCTCTCGTTTGATATGATACACCCCGGAG 780
Db 34754 CTTGACTGGAAGCGCTCGAGCAGGATGTGCTCTCGTTTGATATGATACACCCCGGAG 34813
```

```
QY 781 CGTTATGATAACCGGGATGGCAACTGGGGAGTCTGACGGGGGGATATGACCGGACCTG 840
Db 34814 CGTTATGATAACCGGGATGGCAACTGGGGAGTCTGACGGGGGGATATGACCGGACCTG 34873
```

```
QY 841 CGCTATGACGAAAACAAAATTTTCAGTGGCTATGATCATCTTTCCCTTCGGAACTGG 900
Db 34874 CGCTATGACGAAAACAAAATTTTCAGTGGCTATGATCATCTTTCCCTTCGGAACTGG 34933
```

```
QY 901 AATCGTATCTGAACTGGAAACGAGACAGAAAATAAAGTCTGTGAGCTTGTAGCGAGTGTA 960
```


QY	1186	ACGGATGCATTGGCTGACTGCGGACGCGCTATGAACATCATGACCAATTCGGGGA	1245
Db	8746	CTGGACTCACTGGCATTAAACCTATATGGCGCAGTTATGAATATCAAACTACGTTTGGCGGC	8805
QY	1246	CACATTCAGTCGCGCTGCATATCTGGTCTGGGATGTGGCAGATGCTCGAGCTGAAGGC	1305
Db	8806	CATATTAGCCACAGAGCTATCTGGTGGGACCCACAGATTAACCTGACGGTGAAGGG	8865
QY	1306	GGTGTGACCCAGGATATAGGACCCAGAGATGGGAGCTACATAAAGGATTTAGTGGT	1365
Db	8866	GGTGTGACCGGGATATAAACCCTACCTAGCCAGTTGCATATGGGATTTAGCGGT	8925
QY	1366	GTGTCCGGGAGGAAACAAATCTACTTGTGAACCCCGACCTGAAGCCGGAAGAGAC	1425
Db	8926	GTCAAGGCCAAGCACCACATCACGACGATTTGGTAACCTTAAGCTTGAACCCAGAGTAGT	8985
QY	1426	GTCAATTTATGAGGTGGGTGTATTACGATAACCCCGCGTCTGAATGCCATGTGACA	1485
Db	8986	GTCAACACCGAGTGGGTGTCTATTATGAACGAAACGAGGATTTGGCGTAAATGTGACG	9045
QY	1486	GGTTTATGACTGCTCTCCAAACAGATTTCTCTTTATTCATAAATGATACACCAAT	1545
Db	9046	CTGTTCACACCGCTTCAGGAACAAATCAACAGCGTTTCGATTTGATATACCAACAGC	9105
QY	1546	AGCTATGTAAACAGCGGAGGCGGCTGCAGGCTGTGAATTTGCCGCGCATTTGCCG	1605
Db	9106	ACCTACACCAAGTTCGTAAGCGATCACGCGGATTTAGGTCGATCCACCATACCA	9165
QY	1606	CTGTGGTCAGAGGATGTCACTGCTGTCTCACTGAATTAACCTTGGACCCGAAAGTGAACACGT	1665
Db	9166	CTCTGGTCTGACGATTTGATGCTGGGATATAAATACACCTTTACCGATGTGACAGAA	9225
QY	1666	GATGGTATACAAAGTGGCGCGCTGAGTTATACCCCTGAACACATGTTGAATGGGAA	1725
Db	9226	GATGGCAACATAAAGGGCTCGCTGACGAATACACCGGAAATATGTTTAAATGCCCG	9285
QY	1726	CTGAACCTGGCAGATCACCGAAGGTGGCATCATGCTGGTGGTCCCGTTATCGCGGAA	1785
Db	9286	CTGAACCTGGATATCAACGAGCAGTTGTCCACTTGTGTAAGCGGAGTATGCGATAA	9345
QY	1786	ACACCGATTTCCACGAGATTTATCGTCACTGAGCGGTGTACAGAAAGATGTATGAT	1845
Db	9346	ACGGCAGCTTTTACCCAAAATACGGCAATTTGTCTGCGGCAACAAAGTGTTTACAC	9405
QY	1846	GAGAAAGGAGATACCTGAAGCCTGACGGTGGTGGATCGAGTCTGCTGCGGAGATG	1905
Db	9406	AATTTAGCTCAGAGTTTAACTTTAGCGTACTTAAATTTGGGGTGGCTTATAGGTC	9465
QY	1906	ACGATGCCCTGACGCTGAAATGCTGCGGTGAATAACCTCTCAACAAAGATTTACAGTAC	1965
Db	9466	ACTAAGGATGTGACACTGAATGTGCGAGTCAATAACCTGTTGATAAAGATTTACTCGC	9525
QY	1966	GTGAGCTGTACAGTCCGCTAAGATGCTGTATGCGGTTGATTTCTTCAGACGGGA	2025
Db	9526	ACCCACATTTTCTGTGGGAGACGCCACCACTACTGCGGGTGAATTTACTTCACTCGCT	9585
QY	2026	TCATCAACACAGGATGTGATACCTGAGCGAAATTTACTGGATGTCGCTGAATCATCAG	2085
Db	9586	CAAAGACTGCCGGTTATGTTGATCTGCTGCGGTAATTTACTGCGGATTCGCTGATTTAT	9645
QY	2086	TTCT 2089	
Db	9646	TTCT 9649	

RESULT 15

AJ414156/C

LOCUS

DEFINITION

AJ414156 Yersinia pestis strain CO92 complete genome; segment 16/20.

ACCESSION

AJ414156.1

VERSION

AJ414156.1

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

JOURNAL

COMMENT

FEATURES

source

gene

CDS

misc_feature

misc_feature

misc_feature

gene

CDS

Yersinia pestis CO92
Yersinia pestis CO92
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Yersinia.

1 (bases 1 to 220050)
Parkhill, J., Wren, B.W., Thomson, N.R., Titball, R.W., Holden, M.T.G.,
Prentice, M.B., Sebahia, M., James, K.D., Churcher, C., Mungall, K.L.,
Baker, S., Basham, D., Bentley, S.D., Brooks, K., Cerdano-Tarraga, A.M.,
Chillingworth, T., Cronin, A., Davies, R.M., Davis, P., Dougan, G.,
Fellwell, T., Hanlin, N., Holroyd, S., Jagels, K., Leather, S.,
Karlyshev, A.V., Moule, S., Oyston, P.C.F., Quail, M., Rutherford, K.,
Simmonds, M., Skelton, J., Stevens, K., Whitehead, S., and Barrall, B.G.
Genome sequence of Yersinia pestis, the causative agent of plague
Nature 413 (6855), 523-527 (2001)

2 (bases 1 to 220050)
Parkhill, J.
Direct Submission
Submitted (04-OCT-2001) Submitted on behalf of the Yersinia
sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk

Notes:
Details of Y. pestis sequencing at the Sanger Centre are available
on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/Y_pestis/).

Location/Qualifiers
1. 220050

/organism="Yersinia pestis CO92"
/mol_type="genomic DNA"
/strain="CO92"
/db_xref="taxon:214092"
/note="biovar: Orientalis"
complement(191..715)
/gene="arol"
/note="synonym: YPO3215"
complement(191..715)
/gene="arol"
/EC_number="2.7.1.71"
/note="Similar to Escherichia coli shikimate kinase II
AroL SW:AROL_ECOLI (P08329) (174 aa) fasta scores: E():
1.5e-31, 55.4% id in 168 aa, and to Erwinia chrysanthemi
shikimate kinase AroL or AroM SW:AROL_ERWCH (P10880) (173
aa) fasta scores: E(): 0, 63.7% id in 168 aa"

/codon_start=1
/transl_table=11
/product="shikimate kinase II"
/protein_id="CAC92450.1"
/db_xref="GI:15981151"
/db_xref="GOA:Q8ZC15"
/db_xref="SWISS-PROT:Q8ZC15"
/translation="MTQTFMGARGAGKTTIGKALQAQALGYFVDTDLFMQOTSMT
VASVSESGDGRFLRESMALQVATPKVTATGGAVLSSENRAFMRDHRVYDRA
SAVLAKRLAADPEEAQPSLTGKPIVEILDVLASREALYQDVAAHVLDGTQPTSLV
VEQLQMLTCEMVK"
complement(209..709)
/gene="arol"

/note="Pfam match to entry PF01202 SKI, Shikimate kinase,
score 245.00, E-value 1.1e-69"
complement(470..544)
/gene="arol"
/note="P801128 Shikimate kinase signature."
complement(668..691)
/gene="arol"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)."
complement(962..1519)
/gene="YPO3216"
complement(962..1519)
/note="Similar to Rhizobium meliloti phospholipid
N-methyltransferase PmtA TR:AAG10237 (EMBL:AF201699) (200
aa) fasta scores: E(): 1.1e-07, 25.9% id in 185 aa, and to
Bradyrhizobium japonicum PmtA protein TR:Q9LCT2

misc_feature
/note="Pfam match to entry PF01202 SKI, Shikimate kinase,
score 245.00, E-value 1.1e-69"
complement(470..544)
/gene="arol"

misc_feature
/note="P801128 Shikimate kinase signature."
complement(668..691)
/gene="arol"

misc_feature
/note="PS00017 ATP/GTP-binding site motif A (P-loop)."
complement(962..1519)
/gene="YPO3216"
complement(962..1519)
/note="Similar to Rhizobium meliloti phospholipid
N-methyltransferase PmtA TR:AAG10237 (EMBL:AF201699) (200
aa) fasta scores: E(): 1.1e-07, 25.9% id in 185 aa, and to
Bradyrhizobium japonicum PmtA protein TR:Q9LCT2

gene
complement(962..1519)
CDS
/note="Similar to Rhizobium meliloti phospholipid
N-methyltransferase PmtA TR:AAG10237 (EMBL:AF201699) (200
aa) fasta scores: E(): 1.1e-07, 25.9% id in 185 aa, and to
Bradyrhizobium japonicum PmtA protein TR:Q9LCT2

QY 358 AATACCGGTTTCATGCCCTCTCTGCGCCCATTCAGCGTATTCAGGTTATTCAGGGGGCCG 417
 Db 147953 AATACCGGTTTCATGCCCTCTCTGCGCCCATTCAGCGTATTCAGGTTATTCAGGGGGCCG 147894
 QY 418 ATGTCCACACTGTATGGCTCTGATCGATGGCGGTGTGTGATATATCATTCACAGAAAG 477
 Db 147993 ATGTCCACACTGTATGGCTCTGATCGATGGCGGTGTGTGATATATCATTCACAGAAAG 147834
 QY 478 AATCAGACAAATGGCTCTCTTCGGTCAATGCAGGGCTGAATCTGCAGGAAAGCAAAA 537
 Db 147833 ATCACCAGCGTGGCCAGCTCAGCAACGCTTGAAACACACCTTCACAGGAAATACGGCT 147774
 QY 538 TGGGTTACAGACGCCAGTTTAATTTCTGAGCAGTGGTCCCTTGTGGATGATTCGTG 597
 Db 147773 TTTGGTGACAGTTCAAAAGTTCTCTTTTACTCCAGTGGCGCAGCGTTGAAGATCAACTG 147714
 QY 598 AGCTGCGAGGTACCGGTAGCACACACAGCTCAGGTTTCATCGGTCAATC-----A 651
 Db 147713 GGGCTGTGTTACGTGCACGATTTTCCGGCGGATGCCCTCTCGGTTGAATCCTCCAAT 147654
 QY 652 CTGAGCGATACAGACGCCAGCTTATTCCTTATCCACGAGTACAGAAATTAATCTT 711
 Db 147653 ACGGGAGTAGAATTAAGCCGTGTGGTCCCAATCCAGTAAAGCAGACATATAATCTC 147594
 QY 712 GGTGACGCTCTGACTGGAAGGGTCCGACAGGATGTCTCTGTTTGNATGGATACC 771
 Db 147593 GGTGGCAACTATTTTGGCAATTTGAATACCAGCACTTTGGTGGATGGATATTT 147534
 QY 772 ACCGGCAGCGTTATGATAACCGGGATGGCAACTGGGA-----GTCTGACGGGGGA 825
 Db 147533 GCTAACCAAAATATGATACAGTGCACCACTAGGTACCATTTGTCACGAGCGGT 147474
 QY 826 TATGACCGGACCTTGGCTATGAGCGAACAATAATTTACGTGGCTATGATCATCTTC 885
 Db 147473 TACGAAGATCTCTGGCTTATCAACCGCGTAAATAACCTTGGGCAACGATAATCGTTG 147414
 QY 886 ACCTTCGGAACATGGAATCGTATCTGAACTGGAACGAGACAGAAATAAAGTCTGTAG 945
 Db 147413 GACTTTGGTACCTTGGAAATTCGAGCTTATCTTATACCAACTGAAATATTTGGCCG---C 147357
 QY 946 CTGTGACGAGTGTACTGAAGCCGCAAAATGGGGCTTTCGGGTGAGCGCGGAGCTT 1005
 Db 147356 TTAATCACCAATCATCCGTGCGCCCAAGGCTCTGGGTGGCAGGTGAAAAAGTCTGTG 147297
 QY 1006 AAGGAATCGAACCTTATCTTGAATTCATTTACTGCTTACCCCTCTGGGCAATCTCATCTG 1065
 Db 147296 AAAATACCAATATCATCTTGTATAGCAAGCTGTCGCCCCGTGGTGTATTAACCAATG 147237
 QY 1066 GTTACGGTGGGGCGAGTTTCAGAGCTCGTCCATGAAAGACGAGTTTCTTGGCCAGC 1125
 Db 147236 GTGACCTTTGGCGTCAATATTTGGAATGCCATCATGAAAGCGGCATCTACTGCAAA 147177
 QY 1126 ACAGGTGAACTTTCCGGCAGAAAGCTGTCTGGTATTTGCTGAGATGAGTGGCATCTC 1185
 Db 147176 AATGGCGATGAGTTTGCACAAAGATGCTGTCTATTAATCAGGGAAGATGGAATGGCGTCTG 147117
 QY 1186 ACGGATGCACTTTCGGTGTGCTGGGCGAGCGCTATGAACATCATGCAATTCGGGGGA 1245
 Db 147116 CTGGACTCACTGSCATTAACTTATGCGCAGCTTATGATATCAAACTACGTTTGGCGG 147057
 QY 1246 CACTTCAGTCCGGGTGCATCTCTGTCTGGATGTGGGAGTGCCTTGGACGCTGAAAGC 1305
 Db 147056 CATATTAGCCACAGGCTTATCTGTGTGGGCGCACAGGATAAATGGAACGCTGAAAGG 146997
 QY 1306 GGTGTACCAACGGGATATAAGGCACCAAGATGGGCGACTACATAAAGGGATTTAGTGT 1365
 Db 146996 GGTGTACGACGCGGTATATAAACCCCACTTAGCCAGTTGCATATGCGATTAGCGGT 146937
 QY 1366 GTGTCGGGCGAGGAAAAACAATCTATTGTTGAACCCGACCTGAAGCCGGAAGAGC 1425
 Db 146936 GTCACAGGCCAAGGCCACCATCAGACGATTGGTAACCCCTAAGCTTGAACCAAGAGTAGT 146877

QY 1426 GTCACTTATGAGCTGGGGTGTATTACGATAACCCCGCGCTGCAATGCAATGTCACA 1485
 Db 146876 GTCAACACCGAGTGGGTGTCTATTATGAAACGAAACGGGATTTGGCGCTAATGTGACG 146817
 QY 1486 GGTTTTATGACTGACTTCTCCAAAGATTTCTCTTATTCATTAATGATTAACCAAT 1545
 Db 146816 CTGTTCCAAACCGCTTCAGGAACAAATCAACACGCTTCGATGATTAATACCACG 146757
 QY 1546 AGCTATGTAACACAGCGGAAAGCCCGCTTGCACGGTGTGGAATTTGCCGACATTCGCG 1605
 Db 146756 ACCTACACCAACGTCGTAAGCGATCACGTAGGTTAGTTCGATCCACCATACCA 146697
 QY 1606 CTGTGCTCAGAGGATGTCCAGCTGTCTGATTAACCTGGAACCGGAGTGAACACGT 1665
 Db 146696 CTCTGCTCAGATTTGATGCTGGGAATAATACACCTTTACCGATAGTGAACAGAAA 146637
 QY 1666 GATGTGATAACAAAGGTGCGCCCTGAGTTATACCCCTGAACACATGGTGAATGCGAAA 1725
 Db 146636 GATGGCAACAATAAGGGCTCGCTGACGATTAACCGAAATAATGGTTAATGCCCG 146577
 QY 1726 CTGAACCTGGCAGATCACCGAAGAGTGGCATCATGGTGGGTGCCCTTATCGCGGAAA 1785
 Db 146576 CTGAACCTGGAATATCAACGAGCAGTTGTCCACTTTGGTTGAAAGCGGAGTATCGCAGTAA 146517
 QY 1786 ACACACGTTTTCACCCAGATTTATTCGTCTACTGAGCGCTGTACAGAAAGTGTATGAT 1845
 Db 146516 ACGGACAGTTTATCCCAAAATTCGCGAAATTTGTCTCGGCAACAAAGTGGTTTACAAC 146457
 QY 1846 GAGAAAGGAGAAATACCTGAAAGCCTGGAAGCTGGTGGATGAGTCTGTCTGGAAGATG 1905
 Db 146456 AATTTAGGCTCAGAGTTTAAACCTTTTAGCGTACTTAATTTGGGGTGGCTTAAAGTC 146397
 QY 1906 ACGGATGCCCTGACGCTGAATGCTGCGTGAATAACCTGCTCAACAGGATTAACAGTAC 1965
 Db 146396 ACTAAGGATGTGACACTGATGCTGATGCTCAATAACCTGTTGGATAAAGATTTTACTCGC 146337
 QY 1966 GTGAGCCTGTACAGTGGCGGTGAAGATGAGTACGCTGTATGCGCGTGTACTTCCAGACGGGA 2025
 Db 146336 ACCACATTTTGTCTGTGGGACGGCACCACTACTGCGGGTGTACTTCACTCGTCT 146277
 QY 2026 TCATCAACACAGGATATGTATACCTGAGGAAATTAATCTGATGTGCTGAACTATCAG 2085
 Db 146276 CAAAGCACTGCGGTTATGTTGTAACCTGCGGCTAATTAATCTGCGTATCCGTAATGTAAT 146217
 QY 2086 TTCT 2089
 Db 146216 TTCT 146213

Search completed: October 12, 2004, 15:56:10
 Job time : 8126 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: October 13, 2004, 11:41:17 : Search time 131 Seconds
(without alignments)

10072.497 Million cell updates/sec

Title: US-10-625-972-4

Perfect score: 3732

Sequence: 1 atcggaataaccactctggc.....cgctgaactatcagttctga 2091

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DEV=xlh
-O/cn2 1/USPTO spool/US10625972/runat 12102004 130105 26819/app query.fasta_1.2247
-DB=STREMBL_25 -QFWT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40 cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10625972 @CGN 1 1 177 @runat 12102004 130105 26819 -NCPU=6 -CPU=3
-NO MAP -LARGEQUERY -NEG SCORE=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DRV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3632	97.3	696	16 Q9LAP1	Q9lap1 escherichia

RESULT 1

ID	Q9LAP1	PRELIMINARY;	PRT;	696 AA.
AC	Q9LAP1	Q8X9P7;		
DT	01-OCT-2000	(TRENBLrel. 15, Created)		
DT	01-JUN-2002	(TRENBLrel. 21, Last sequence update)		
DT	01-JUN-2003	(TRENBLrel. 21, Last annotation update)		
DE	Adhesin (Putative receptor) (Iha adhesin).			
GN	IHA OR Z1178 OR Z1617 OR ECS1360.			
OS	Escherichia coli O157:H7.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Escherichia.			
OX	NCBI_TaxID=83334;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=86-24;			
RX	MEDLINE=20143752; PubMed=10678953;			
RA	Tarr P.I., Bilge S.S., Vary J.C. Jr., Jelacic S., Habeeb R.L.,			
RA	Ward T.R., Baylor M.R., Besser T.E.;			
RT	"Iha: a novel Escherichia coli O157:H7 adherence-conferring molecule			
RT	encoded on a recently acquired chromosomal island of conserved			
RT	structure."			
RL	Infect. Immun. 68:1400-1407(2000).			
RN	[2]			
RP	SEQUENCE FROM N.A.			

ALIGNMENTS

Q8fdx2 escherichia
Q93se1 escherichia
Q87518 escherichia
Q8v192 yersinia pe
Q8v168 escherichia
P70847 bordetella
Q9pgg3 campylobact
Q7vi87 helicobacte
O07651 campylobact
Q87g14 yersinia pe
Q87lk9 vibrio para
Q93rd3 agrobacteri
Q8v144 escherichia
Q8faq7 escherichia
Q93fb0 escherichia
Q88x7 shewanella
Q889g6 pseudomonas
Q82597 salmonella
Q82n10 salmonella
Q8x633 escherichia
Q8cvw8 escherichia
Q91212 pseudomonas
Q840d4 acinetobact
Q88kp8 pseudomonas
Q7vv12 bordetella
Q7w19 bordetella
Q7w108 bordetella
Q8cwal escherichia
Q45341 bordetella
Q8xbw7 escherichia
Q821b8 shigella fl
Q9rq19 escherichia
Q8cy02 escherichia
Q8pj6 xanthomonas
Q93k73 escherichia
Q82818 salmonella
Q68598 pseudomonas
Q91527 pseudomonas
Q8gh18 escherichia
Q8z40 salmonella
Q8z4h1 salmonella
Q8rmh4 salmonella
Q8zmn0 salmonella

RC STRAIN=86-24;
RA Tarr P.I., Bilge S.S., Vary J.C., Jelacic S., Habeeb R.L., Ward T.R.,
RA Baylor M.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RP [3]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck B.J., Davis N.W., Lim A., Dimalanta E.T., Potamoukis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
RL Nature 409:529-533(2001).
RP [4]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 050952;
RX MEDLINE=2115231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Otsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12";
RL DNA Res. 8:11-22(2001).
DR ENBL; AF126104; AAF36432.2; -;
DR ENBL; AE005310; AAG55732.1; -;
DR ENBL; AE005274; AAG55323.1; -;
DR ENBL; AP002554; BAB34783.1; -;
DR PIR; G85607; G85607.
DR PIR; H90798; H90798.
DR HSP; P05825; IFEF.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0005215; P:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR000531; TonB boxC.
DR Pfam; PF00593; TonB dep Rec; 1.
KW Receptor; Complete proteome.
SQ SEQUENCE 696 AA; 76512 MW; 7651E329169D9064 CRC64;

Alignment Scores:
Pred. No.: 3,18e-223 Length: 696
Score: 3632.00 Matches: 695
Percent Similarity: 99.86% Conservative: 0
Best Local Similarity: 99.86% Mismatches: 1
Query Match: 97.32% Indels: 0
DB: 16 Gaps: 0

US-10-625-972-4 (1-2091) x Q9LAP1 (1-696)

QY 1 ATCCGATACCACTCTGCTCCGTCAGTATCCCTGTCCTGGATTTTCACCCAGCAGC 60
Db 1 MetArgilethrThrLeuAlaSerValValIleProCysLeuGlyPheSerAlaSerSer 20
QY 61 ATAGCTGCTCCAGAGGATGTGATGTCTCGCATCCGGCTATCAGAAAAAGCTGACT 120
Db 21 IleAlaAlaAlaGluAspValMetIleValSerAlaSerGlyTyThrGluLysIleLeuThr 40
QY 121 AACGAGCCGCCAGTCTTCTGTGATACAGGAGGAATTCAGTCCAGCCAGTACCAAC 180
Db 41 AsnAlaAlaAlaSerValSerValIleSerGlnGluLeuGlnSerSerGlnTyHis 60
QY 181 GATCTCCGCGAGGCTCTGAGATCAGTAGAGGCTGTGGATGTTGAAAGTGGTACGGGTAAA 240
Db 61 AspLeuAlaGluAlaLeuArgSerValGluGlyValAspValGluSerGlyThrGlyLys 80
QY 241 ACCGAGGCTGGAATCAGATCCAGGAATGCCAGGAGTACACGCTGATCTGATT 300
Db 81 ThrGlyGlyLeuGluIleSerIleArgGlyMetProAlaSerTyThrLeuIleLeuIle 100

QY 301 GATGTTGTTCTGTCAGGCGGGAAGCAGTGCAGTCCCAACGGTTTTTCTGCCATGAAT 360
Db 101 AspGlyValArgGlnGlySerSerAspValThrProAsnGlyPheSerAlaMetAsn 120
QY 361 ACCGGTTTCATGCCCTCTCGCCGCAATTCAGGCTATTGAGGTATTCAGGGGCGCATG 420
Db 121 ThrGlyPheMetProProLeuAlaAlaIleGluArgIleGluValIleArgGlyProMet 140
QY 421 TCCACACTGTATGGCTCTTCATGTCGATGGCGCGTCTGGTGAATATCATACAGAAAGAAT 480
Db 141 SerThrLeuTyThrGlySerAspAlaMetGlyGlyValValAsnIleIleThrArgLysAsn 160
QY 481 GCAGCAAAATGGCTCTCTTCCTCAATGCAGGCTGAATCTGCAGGAAAGCAACAATGG 540
Db 161 AlaAspLysTrpLeuSerSerValAsnAlaGlyLeuAsnLeuGlnGluSerAsnLysTrp 180
QY 541 GGTAAACAGCAGCCAGTTTAAATTTCTGGAGCAGTGCCTCCCTTGTGGATGATTTCTGTCCAGC 600
Db 181 GlyAsnSerSerGlnPheAsnPheTrpSerSerGlyProLeuValAspSerValSer 200
QY 601 CTGCAAGTACCGGTAGCACACACACGCTCAGCGTTTCATCGGTTCACATCAGTACGAT 660
Db 201 LeuGlnValArgGlySerThrGlnGlnArgGlnGlySerSerValThrSerLeuSerAsp 220
QY 661 ACAGCAGCAGCGCTATTCCTTATCCACGAGTGCACAGATTAATATCTGTGTGACCT 720
Db 221 ThrAlaAlaThrArgIleProTyProThrGluSerGlnAsnTyThrAsnLeuGlyAlaArg 240
QY 721 CTTGACTCGAAGCGCTCGGACGAGTGTGCTCTGTTTTCATATGATATGATACACCCGCGAG 780
Db 241 LeuAspTrpLysAlaSerGluGlnAspValLeuTrpPheAspMetAspThrThrArgGln 260
QY 781 CGTTATGATAACCGGATGGCAACTCGGAGTCTGACGGGGGATATGACCGAACCTG 840
Db 261 ArgTyThrAspAsnArgAspGlyGlnLeuGlySerLeuThrGlyTyThrAspArgThrLeu 280
QY 841 CGCTATGAGCAAAACAAAATTTTCAGCTGCTATGATCATATCTTTCACCTTCGGAACATGG 900
Db 281 ArgTyThrGluArgAsnLysIleSerAlaGlyTyThrAspHisThrPheThrGlyThrTrp 300
QY 901 AAATCGTATCTGAATCGAAACGAGACAGAAATAAAGTCTGCTGAGTGTGTACCGAGTGA 960
Db 301 LysSerTyThrLeuAsnTrpAsnGluThrGluAsnLysGlyArgGluLeuValArgSerVal 320
QY 961 CTGAAGCGCAGCAAAATCGGCGCTTCGCGTCAGCCGCGGAGCTTAAGGAATCGAACCTT 1020
Db 321 LeuLysArgAspLysTrpGlyLeuAlaGlyGlnProArgGluLeuLysGluSerAsnLeu 340
QY 1021 ATCTGAAATTCATTACTGCTTACCCCTCTGGGAGAATCTCATCTGTTACGTTGGGGGGC 1080
Db 341 IleLeuAsnSerLeuLeuLeuThrProLeuGlyGluSerHisLeuValThrValGlyGly 360
QY 1081 GAGTTTCAGAGCTCGTCCATGAAGCGAGTGTCTCTTCCGAGCAGCAGGTGAACCTTTC 1140
Db 361 GluPheGlnSerSerSerMetLysAspGlyValValLeuAlaSerThrGlyGluThrPhe 380
QY 1141 CGGCAGAAAAAGCTCGTCCGTTATTTGTCAGGATGAGTGGCATCTCAGGATGACATTGGC 1200
Db 381 ArgGlnLysSerTrpSerValPheAlaGluAspGluTrpHisLeuThrAspAlaLeuAla 400
QY 1201 CTGACTCGGGGAGCGCTATGAACATCATGACCAATTCGGGGGACACTTCAGTCCGCGT 1260
Db 401 LeuThrAlaGlySerArgTyThrGluHisHisGluGlnPheGlyGlyHisPheSerProArg 420
QY 1261 GCATATCTGCTCGGATGTGGCAGATGCTGACGCTGAAAGCGGCTGTGACCAACCGGA 1320
Db 421 AlaTyThrValTrpAspValAlaAspAlaTrpThrLeuLysGlyGlyValThrThrGly 440
QY 1321 TATAAGGCCCCAGAAATGGGGCAGCTACATAAAGGATTAGTGGTGTGTCGGGGCAGGGA 1380
Db 441 TyrLysAlaProArgMetGlyGlnLeuHisLysGlyIleSerGlyValSerGlyGlnGly 460
QY 1381 AAAACAAATCTACTTGGTAAACCCCGACCTGAAGCCGGAAGAGAGCGTTCAGTTATGAGGCT 1440

DR GO: 0004872; F:receptor activity; IEA.
DR GO: 0005215; F:transporter activity; IEA.
DR GO: 0006810; P:transport; IEA.
DR InterPro: IPR000531; TonB_boxC.
DR Pfam: PF00593; TonB_dep_Rec; 1.
KW Receptor; Complete Proteome;
SQ SEQUENCE 696 AA; 76511 MW; F96D10B59C553CE8 CRC64;

Alignment Scores: 6.62e-223 Length: 696
Pred. No.: 3627.00 Matches: 694
Score: 3627.00
Percent Similarity: 99.86% Conservative: 1
Best Local Similarity: 99.71% Mismatches: 1
Query Match: 97.19% Indels: 0
DB: 16 Gaps: 0

US-10-625-972-4 (1-2091) x Q8FDX2 (1-696)

Qy 1 ATCGAATAACCACTCTCGGCTTCGGTAGTCATCCCTCTCGGATTTTCAGCCAGCAGC 60
Db 1 MetArgIleThrThrLeuAlaSerValValIleProCysLeuGlyPheSerAlaSerSer 20
Qy 61 ATAGCTCTCCAGAGGATGTGATGTCTCGGCATCCGGCTATGAGAAAAGCTGACT 120
Db 21 IleAlaAlaAlaGluAspValMetIleValSerAlaSerGlyTyrGluLysLysLeuThr 40
Qy 121 AACCCAGCCCGCTGCTGCTGATGTAGCAGGAGGATTCAGTCCAGCCAGTACCAC 180
Db 41 AsnAlaAlaAlaSerValSerValIleSerGluGluLeuGlnSerSerGlnTyrHis 60
Qy 181 GATCGCGGAGGCTCTGAGATCAGTAGACGGTGTGATCTTGAAGTGTAGCGGTAAA 240
Db 61 AspLeuAlaGluAlaLeuArgSerValGluGlyValAspValGluSerGlyThrGlyLys 80
Qy 241 ACCGAGGGTGGAAATCAGATCCGAGGAGGATCCAGCCAGTTCACGCTGATCTGATT 300
Db 81 ThrGlyGlyLeuGluIleSerIleArgGlyMetProAlaSerTyrThrLeuIleLeuIle 100
Qy 301 GATCGTGTCTGCTAGCGCGGAGGAGCAGTGACGTACTCCCAACGGTTTTCTGCCATGAAT 360
Db 101 AspGlyValArgGlnGlyGlySerSerAspValThrProAsnGlyPheSerAlaMetAsn 120
Qy 361 ACCGGTTCATGCCCCCTCTGGCCCGCATTCAGCGGTATTCAGGTATTCAGGGGCCATG 420
Db 121 ThrGlyPheMetProProLeuAlaAlaIleGluArgIleGluValIleArgGlyProMet 140
Qy 421 TCCACACTGTATGCTCTGATCGCATGGCGGTGTGTAATATCATTCACAGAAAAGAT 480
Db 141 SerThrLeuTyrGlySerAspAlaMetGlyValValAsnIleIleThrArgLysAsn 160
Qy 481 GCAGACAAATGGCTCTCTTCCTCAATCCAGGGGTGAATTCGAGGAAAGCAACAAATGG 540
Db 161 AlaAspLysTyrLeuSerSerValAsnAlaGlyLeuAsnLeuGlnGluSerAsnLysTrp 180
Qy 541 GGTAAACAGCAGCCAGTTTAATTTCTGGAGCAGTGGTCCCTCTGGTGGATTCCTGTGAGC 600
Db 181 GlyAsnSerSerGlnPheAsnPheTrpSerSerGlyProLeuValAspSerValSer 200
Qy 601 CTGCAAGTACCGGTAGCACACAAACAGCGTTCAGGGTTCATCGGTACATCATCTAGCGAAT 660
Db 201 LeuGlnValArgGlySerThrGlnGlnArgGlnGlySerSerValThrSerLeuSerAsp 220
Qy 661 ACAGCAGCAGCGGTATTCCTTATCCACGAGGATCACAGATTTATATCTTGTGTCAGGT 720
Db 221 ThrAlaAlaThrArgIleProTyrProThrGluSerGlnAsnTyrAsnLeuGlyAlaArg 240
Qy 721 CTGAGTCTGGAAGGCTCGGAGCAGATGTCTCTGTTGTATGTATGATGATACACCCGGCAG 780
Db 241 LeuAspTyrLysAlaSerGluGlnAspValLeuTrpPheAspMetAspThrThrArgGln 260
Qy 781 CGTTATGATAACCGGGATGGGCAACTGGGAGTCTGACGGGGGATATGACCGGACCTG 840
Db 261 ArgTyrAspAsnArgAspGlyGlnLeuGlySerLeuThrGlyGlyTyrAspArgThrLeu 280

Db 461 LysThrAsnLeuLeuGlyAsnProAspLeuLysProGluGluSerValSerTyrGluAla 480
Qy 1441 GGGGTGTATACGATAACCCCGCGCTGATGATCCCAATGTCACAGGTTTATGACTCAC 1500
Db 481 GlyValTyrTyrAspAsnProAlaGlyLeuAsnAlaAsnValThrGlyPheMetThrAsp 500
Qy 1501 TTCTCCAAACAGATTGCTCTTATTCCTAATAATGATAACACCAATAGCTATGTAACAGC 1560
Db 501 PheSerAsnLysIleValSerTyrSerIleAsnAspAsnThrAsnSerTyrValAsnSer 520
Qy 1561 GGAAGAGCCCGGTTCACCGGTGTGAATTTGCCGGCACAATTGCCCTGTGTCTGACAGGAT 1620
Db 521 GlyLysAlaArgLeuHisGlyValGluPheAlaGlyThrLeuProLeuTrpSerGluAsp 540
Qy 1621 CTCACGCTGCTACTGAATTACACCTGGACCCGAGTGAACACGCTGATGTGTAACAAA 1680
Db 541 ValThrLeuSerLeuAsnTyrThrThrArgSerGluGlnArgAspGlyAspAsnLys 560
Qy 1681 GGTGGCCGCTGAGTTATACCCCTGAACACATGTGTAATGCGAACTGAACCTGGCAGATC 1740
Db 561 GlyAlaProLeuSerTyrThrProGluHisMetValAsnAlaLysLeuAsnTrpGlnIle 580
Qy 1741 ACCGAGAGCTGCATCATGCTGGTGGTGGTCCGCTTATCGGGGAAACACACGCTTTCACC 1800
Db 581 ThrGluGluValAlaSerTrpLeuGlyAlaArgTyrArgGlyLysThrProArgPheThr 600
Qy 1801 CAGAATTATTCCTCACTGAGCGCTGTACAGAAAGAAAGTGTATGATGAGAAAGGAGATAC 1860
Db 601 GlnAsnTyrSerSerLeuSerAlaValGlnLysLysValTyrAspGluLysGlyGluTyr 620
Qy 1861 CTGAAAGCTGACCGGTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1920
Db 621 LeuLysAlaTrpThrValValAspAlaGlyLeuSerTrpLysMetThrAspAlaLeuThr 640
Qy 1921 CTGAAAGCTGCGGTGAATTAACCTGCTCAACAGGATTCAGTACGCTGAGCCTGATACAGT 1980
Db 641 LeuAsnAlaAlaValAsnAsnLeuLeuAsnLysAspTyrSerAspValSerLeuTyrSer 660
Qy 1981 GCGGTGAAGAGTACGCTGTATGCGCGGTGATTAATTCACAGCGGATCATCAACACAGGA 2040
Db 661 AlaGlyLysSerThrLeuTyrAlaGlyAspTyrPheGlnThrGlySerSerThrThrGly 680
Qy 2041 TATGATATACCTGAGCGAATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 2088
Db 681 TyrValIleProGluArgAsnTyrTrpMetSerLeuAsnTyrGlnPhe 696

RESULT 2
Q8FDX2 PRELIMINARY; PRT; 696 AA.
AC Q8FDX2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative receptor.
GN C3610.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22398224; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
of uropathogenic Escherichia coli."
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
DR EMBL: AE016766; AA082058.1; -;
DR GO: 0015020; C:membrane; IEA.

QY 841 CGCTATGAGCGAAACAAATTTACGCTGGCTATGATCATATCTTTACCTTCGGAACATGG 900
Db 281 ArgTyrGluArgAsnLysIleSerAlaGlyTyrAspHisThrPheThrPheGlyThrTrp 300
QY 901 AAATCGTATCTGAACCTGGACGACAGACAAATAAAGTGTGAGCTGTGACGAGTGA 960
Db 301 LysSerTyrLeuAsnTrpAsnGluThrGluAsnLysGlyArgGluLeuValArgSerVal 320
QY 961 CTGAAGCGCGCAAAATGGGGCTTTCGGGTCAGCGCGGAGCTTAAGGAATCGAACCTT 1020
Db 321 LeuLysArgAspLysTrpGlyLeuAlaGlyGlnProArgGluLeuLysGluSerAsnLeu 340
QY 1021 ATCTCGAATTCATTACTGCTTACCCCTCTGGGAGAACTCATCTGTTACGGTGGGGGC 1080
Db 341 IleLeuAsnSerLeuLeuLeuThrProLeuGlyGluSerHisLeuValThrValGlyGly 360
QY 1081 GAGTTTCAGAGCTCGTCCATGAAGACGGAGTTGTCTTCCCTCCAGCACAGCTGAACCTTC 1140
Db 361 GluPheGlnSerSerMetLysAspGlyValValLeuAlaSerThrGlyGluThrPhe 380
QY 1141 CGGCAAGAAAGCTGGTGGTATTTGTCAGGATGAGTGGCATCTCAGGATGCACTTGGC 1200
Db 381 ArgGlnLysSerTrpSerValPheAlaGluAspGluTrpHisLeuThrAspAlaLeuAla 400
QY 1201 CTGACTCGGGCAGCGCTATGACATCATGACCAATTCGGGGGACACTTCAGTCCGCT 1260
Db 401 LeuThrAlaGlySerArgTyrGluHisGlnGluPheGlyGlyHisPheSerProArg 420
QY 1261 GCATATCTGCTGGGATGGCAGATGCTCGGACGCTGAAGCGGCTGTGACACGGGA 1320
Db 421 AlaTyrLeuValTrpAspValAlaAspAlaTrpThrLeuLysGlyGlyValThrThrGly 440
QY 1321 TATAGGCACCAAGATGGGGGAGCTACATAAAGGATGATGCTGTGCTCCGGGAGGA 1380
Db 441 TyrLysAlaProArgMetGlyGlnLeuHisLysGlyLeuSerGlyValSerGlyGlnGly 460
QY 1381 AAAACAAATCTACTTGTGAACCCGACCTGAAGCCGGAAGAGAGCTGAGTTATGAGCT 1440
Db 461 LysThrAsnLeuLeuGlyAsnProAsnLeuLysProGluSerValSerTyrGluAla 480
QY 1441 GGGGTGATTAACATAACCCCGCGCTGTAATGCCAATGTCAAGTTTATGATGAC 1500
Db 481 GlyValTyrTyrAspAsnProAlaGlyLeuAsnAlaAsnValThrGlyPheMetThrAsp 500
QY 1501 TTCTCCAAAGATGTCTTATTCATAAATGATAACCAATAGCTATGTAAACAGC 1560
Db 501 PheSerAsnLysIleValSerTyrSerIleAsnAspAsnThrAsnSerTyrValAsnSer 520
QY 1561 GGAAGCGCGGTGACCGGTGGAAATTTGCGGCACATTCGCGTGTGGTCAGAGGAT 1620
Db 521 GlyLeuAlaArgLeuHisGlyValGluPheAlaGlyThrLeuProLeuTrpSerGluAsp 540
QY 1621 GTCACGCTGTCATGAATACACCTCGACCCGAGTGAACACGTCATGCTGATACAA 1680
Db 541 ValThrLeuSerLeuAsnTyrThrTrpThrArgSerGluGlnArgAspGlyAspAsnLys 560
QY 1681 GGTGCGCGCTGAGTTATACCCCTGAACACATGTTGAATCGCAAACTGAACCTGGCAGATC 1740
Db 561 GlyAlaProLeuSerTyrThrProGluHisMetValAsnAlaLysLeuAsnTrpGlnIle 580
QY 1741 ACCGAGAGGTGGCATCATGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1800
Db 581 ThrGluGluValAlaSerTrpLeuGlyAlaArgGlyArgGlyLysThrProArgPheThr 600
QY 1801 CAGAATATTCTGCTACGCTGTACAGAGAAAGTGTATGATGAGAAAGCAATAC 1860
Db 601 GlnAsnTyrSerSerLeuSerAlaValGlnLysLysValTyrAspGluLysGlyGluTyr 620
QY 1861 CTGAAGCTGAGCGGTGTGATGAGCTGTGCTGTGGAAGATGACGATGCTCCCTGACG 1920
Db 621 LeuLysAlaTrpThrValValAspAlaGlyLeuSerTrpLysMetThrAspAlaLeuThr 640

QY 1921 CTGAATGCTGGGTGAATACCTGCTCAACAGGATTACAGTGAGCTGAGCTGTACAGT 1980
Db 641 LeuAsnAlaAlaValAsnAsnLeuLeuAsnLysAspTyrSerAspValSerLeuTyrSer 660
QY 1981 GCCGTAAAGTACGCTGTATCCGGTGATTAATCTCCAGACGGGATCATCAACACAGGA 2040
Db 661 AlaGlyLysSerThrLeuTyrAlaGlyAspTyrPheGlnThrGlySerSerThrThrGly 680
QY 2041 TATGTGATACCTGACGCAAAATTAATGAGTGTGGTGAACCTATCAGTTTC 2088
Db 681 TyrValIleProGluArgAsnTyrTrpMetSerLeuAsnTyrGlnPhe 696
RESULT 3
Q93SE1
ID Q93SE1 PRELIMINARY; PRT; 695 AA.
AC Q93SE1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=4797/97;
RA Zhang W.L.; Bielaszewska M.; Schmidt H.; Karch H.;
RT "Characterisation of a novel pathogenicity island present in LEE-
negative Shiga toxin-producing Escherichia coli."
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).
DR EMBL; AJ278144; CAC39288.1; -;
DR GO; GO:0001987; C:outer membrane; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0005215; P:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR000531; TonB boxC.
DR Pfam; PF00593; TonB dep_R2C; 1.
KW Hypothetical protein; Membrane; Outer membrane; Receptor; TonB box.
SQ SEQUENCE 695 AA; 76431 MW; 151B7053262CCEAI CRC64;
Alignment Scores:
Pred. No.: 7.6e-215 Length: 695
Score: 3500.50 Matches: 664
Percent Similarity: 98.42% Conservative: 21
Best Local Similarity: 95.40% Mismatches: 10
Query Match: 93.80% Indels: 1
DB: 2 Gaps: 1
US-10-625-972-4 (1-2091) x Q93SE1 (1-695)
QY 1 ATCGGAATACCACTCTGGCTTCGGTAGTCATTCCTGCTCGGATTTTCAGCCAGCAGC 60
Db 1 MetArgIleThrThrLeuAlaSerValValIleProCysLeuGlyPheSerAlaSerSer 20
QY 61 ATAGCTCTCTCAGAGGATGTGATGATGTCTCGGCATCCGGCTATGAGAAAAAGCTGACT 120
Db 21 Met---AlaAlaGluAspValMetIleValSerAlaSerGlyTyrGluLysLysLeuThr 39
QY 121 AACGACCCCGCAGTGTTCCTGTGATTTAGCCAGAGGAATTCAGTCCAGCAGTACCAC 180
Db 40 AsnAlaAlaAlaSerValSerValIleSerGlnGluGluLeuGlnSerGlnTyrHis 59
QY 181 GATCTGCGGAGCTCTGAGATCAGTAGAGGGTGTGGATGTTGAAAGTGGTACCGGTA 240
Db 60 AspLeuAlaGluAlaLeuArgAlaValGluGlyValAspValGluSerGlyThrGlyLys 79
QY 241 ACCGAGGGCTGAAATCAGCATCCGAGGAAATGCCAGCAGTTACAGCTGATCTGATT 300
Db 80 ThrGlyGlyLeuGluIleSerIleArgGlyMetProAlaSerTyrThrLeuIleLeu 99
QY 301 GATGGTGTCTGTCAGGCGGGAAGCAGTACGTGACTCCCAACGGTGTCTCTGCCATGAAT 360

```

Db      100 AspGlyValAlaGlnGlySerGlyAspValThrProAsnGlyPheSerAlaMetAsn 119
Qy      361 ACCGGTTTCAGCCCTCCCTCCGCGCCATTGAGGTTATCAGGGGCCCGATG 420
Db      120 ThrGlyPheMetProProLeuAlaAlaAlaGluAlaGluValIleArgGlyProMet 139
Qy      421 TCCACATGTATGGCTCTGATCGCATGGCGGTGTGGTGAATATCATACAGAAAGAAT 480
Db      140 SerThrLeuTyrGlySerAspAlaMetGlyGlyValValAsnIleIleThrLysLysAsn 159
Qy      481 GCAGACAAATGGCTCTCTCCGTCATACAGGGCTGTAATCTGCAGGAAGCAACAATGG 540
Db      160 AlaAspLysTyrLeuSerSerValAsnAlaGlyLeuAsnLeuGlnGluSerAsnLysTyr 179
Qy      541 GGTAAACAGCAGCCAGCTTTAAATTTCTGGAGCAGTGTCCCTTGTGGATGATCTCTCAGC 600
Db      180 GlyAsnSerSerGlnPheAsnPheTyrPheSerGlyProLeuValAspAspSerValSer 199
Qy      601 CTCGAGGTACCGCGGTAGACACACACAGCGTCAAGGTTTCATCGGCACATCTCAGCGAT 660
Db      200 LeuGlnIleArgGlySerThrGlnGlnArgGlnGlySerValThrSerLeuSerAsp 219
Qy      661 ACAGCAGGACCGCGTATCTTATCCACGAGTACACAGAAATTAATCTTTGGTGCACGT 720
Db      220 ThrSerAlaThrArgIleProTyrProThrGluSerGlnAsnTyrAsnLeuGlyAlaArg 239
Qy      721 CTTGACTGGAAGCGTCGAGCAGCATGTCTCTGTTGTATGATATGATACACACCGCGAG 780
Db      240 IleAspTyrLysAlaThrGluGlnAspValLeuTyrPheAspMetAspThrArgGln 259
Qy      781 CTTTATGATACCGGGATGGCACTGGGAGTCTGACGGGGGATATGACGGGACCGCTG 840
Db      260 ArgTyrAspAsnGlnAspGlyGlnLeuGlySerLeuThrGlyGlyTyrAspHisThrLeu 279
Qy      841 CCCTATGACGGAACAAATTTTCAGCTGGCTATGATCATCTTTCACCTTCGGAACATGG 900
Db      280 ArgTyrGluArgAsnLysIleSerAlaGlyTyrAspHisThrPheThrPheGlyThrTyr 299
Qy      901 AAATCGTATCTGAATCGAACACAGACAGAAATTAAGTCTGTCGAGCTGTACCGCATGTA 960
Db      300 LysSerTyrLeuAsnTyrPheAsnGlnThrGluAsnLysGlyArgGluLeuValArgSerVal 319
Qy      961 CTGACGCGCACAATATGGGGGTTCGCGGTACGCGCGGAGCTTAAGGAATCGAACCTT 1020
Db      320 LeuLysSerAspLysTyrGlyLeuAlaGlyLysProArgGluLeuLysGluSerAsnLeu 339
Qy      1021 ATCTGTAATCTAATCTGTTACCTCCCTCTGGAGATCTCATCTGTTACGTTGCGGGGC 1080
Db      340 IleLeuAsnSerLeuLeuLeuThrProLeuGlyGluSerHisLeuLeuThrValGlyGly 359
Qy      1081 GAGTTTCAGAGCTCGTCCATGAAGACGAGTGTCTTGGCAGACAGCTGAAACTTTC 1140
Db      360 GluTyrGlnSerSerSerMetLysAspGlyValValLeuAlaSerThrGlyGluThrPhe 379
Qy      1141 CGSCAGAAAAGCTGTGCGTATTTGTGAGGATGATGGCATCTCAACGGATCCACTTGG 1200
Db      380 ArgGlnLysSerTyrPheValPheAlaGluAspGluTyrHisLeuThrAspValLeuAla 399
Qy      1201 CTGACTCGGGGACCGCTATGACATCATGACCATTCGGGGGACACTTCAGTCCGGT 1260
Db      400 LeuThrAlaGlySerArgTyrGluHisHisAspGlnPheGlyGlyHisPheSerProArg 419
Qy      1261 GCATATCTGTCGCGATGTGGCAGATGCTGACGCTGAAAGCGCGTGTGACCAAGGA 1320
Db      420 AlaTyrLeuValTyrPheValThrAspAlaTyrThrLeuLysGlyGlyValThrThrGly 439
Qy      1321 TATAAGCCACCCAGATGGGCGAGCTACATAAAGGATTAAGTGTGTGTCGGCGGAGGA 1380
Db      440 TyrLysAlaProThrMetGlyGlnLeuHisLysGlyIleSerGlyValSerGlyGlnGly 459
Qy      1381 AAACAAATCTACTCTGTAACCCCGACCTGAACCGGAGAGAGCGTCAATGATGAGCT 1440

```

```

Db      460 ArgThrAsnLeuLeuGlyAsnProGluLeuLysProGluGluSerValSerTyrGluAla 479
Qy      1441 GGGGTCTATTACGATAACCCCGCTCTGAATGCCAATGTACAGTTTATGACTGAC 1500
Db      480 GlyValTyrTyrAspAsnLeuSerGlyLeuAsnAlaAsnIleThrGlyPheIleThrAsp 499
Qy      1501 TTCTCCAAACAGATCTCTCTTATTCATAAATGATAACACCAATAGCTATGTAACACAGC 1560
Db      500 PheSerAsnLysIleValSerTyrSerIleAspAspAsnThrAsnSerTyrThrAsnSer 519
Qy      1561 GGAAGCGCCGGTGCACCGTGTGGAATTTGCGGCGACATTCGCGCTGTGCTGAGAGGAT 1620
Db      520 GlyLysAlaArgLeuHisGlyValGluPheAlaGlyThrLeuProLeuTyrPheSerGluAsp 539
Qy      1621 GTACCGTCTCACTGATACACTCGACCCGAGTGAACAGCTGATGCTGTATACAA 1680
Db      540 ValThrLeuSerLeuAsnTyrThrTyrThrGlnSerGluGlnArgAspGlyAspAsnLys 559
Qy      1681 GGTGCGCGCTGAGTTATACCCCTGAACACATCGTGAATCGGAACCTGAACCTGCAGATC 1740
Db      560 GlyAlaProLeuSerTyrThrProGluHisMetValAsnAlaLysLeuAsnTyrGlnIle 579
Qy      1741 ACCGAGAGTGGCATCATGGCTGGTGGTGGTTCGCGGTAATCGGGGAAACACCCAGTTTCACC 1800
Db      580 ThrGluGluValAlaSerTyrLeuGlyAlaArgTyrArgGlyLysThrProArgPheThr 599
Qy      1801 CAGATTATTTCGTCACGTGCTGTACAGAAAGTGTATGATGAGAAAGGAGAAATAC 1860
Db      600 ArgAsnTyrSerSerLeuSerAlaValGlnLysValTyrAspGluLysGlyGluTyr 619
Qy      1861 CTGAAAGCTCGGAGTGGTGGATGAGGTCCTGCTGCGGAAGATGACGAGTGCCTGACG 1920
Db      620 LeuLysAlaThrThrValValAspAlaGlyLeuSerTyrLysValThrAspAlaLeuThr 639
Qy      1921 CTGATCTCGCGTGAATACCTGCTCAACAGGATTCACAGTCACTGACGTGACCTGTACAGT 1980
Db      640 LeuAsnAlaAlaValAsnAsnLeuLysAspTyrSerAspValSerLeuTyrSer 659
Qy      1981 GCGGTGAAGTACGCTGTATGCGGTGATTTACTTCCAGACGGATCATCAACACAGGA 2040
Db      660 AlaAspLysSerThrLeuTyrAlaGlyAspTyrPheGlnThrGlySerSerThrThrGly 679
Qy      2041 TATGTGATCTGACGCGAAATTTACTGATCTGCTGACTCACTCAGTTTC 2088
Db      680 TyrValIleProGluArgAsnTyrTyrMetSerLeuAsnTyrGlnPhe 695

RESULT 4
087518 ID 087518 PRELIMINARY; PRT; 669 AA.
AC 087518;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Exogenous ferric siderophore receptor R4.
GN R4.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN (1) SEQUENCE FROM N.A.
RC STRAIN=CF7073;
RX MEDLINE=97342757; PubMed=9199454;
RA Kao J.S., Stucker D.M., Warren J.W., Mobley H.L.;
RT "Pathogenicity island sequences of pyelonephritogenic Escherichia coli
RL CF7073 are associated with virulent uropathogenic strains."
RN Infect. Immun. 65:2812-2820(1997).
RN (2) SEQUENCE FROM N.A.
RC STRAIN=CF7073;
RX MEDLINE=96380392; PubMed=9712795;
RA Guyer D.M., Kao J.S., Mobley H.L.T.;
RT "Genomic analysis of a pathogenicity island in uropathogenic

```


RT Escherichia coli CF7073: distribution of homologous sequences among
RT isolates from patients with pyelonephritis, cystitis, and catheter
RT associated bacteriuria and from fecal samples.";
RL Infect. Immun. 66:4411-4417(1998).
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).
DR EMBL: AF081285; AAC61730.1; -.
DR HSRP; P03825; IEPF.
DR GO: GO:0019857; C:outer membrane; IEA.
DR GO: GO:0004872; F:receptor activity; IEA.
DR GO: GO:0005215; F:transporter activity; IEA.
DR GO: GO:0006810; P:transport; IEA.
DR InterPro: IPR000531; TonB_boxC.
DR Pfam: PF00593; TonB_dep_Rec; i.
KW Membrane; Outer membrane; Receptor; TonB box.
SQ SEQUENCE 669 AA; 73806 MW; B559751F05B57F46 CRC64;

Alignment Scores:
Pred. No.: 8,12e-215 Length: 669
Score: 3500.00 Matches: 667
Percent Similarity: 99.85% Conservative: 1
Best Local Similarity: 99.70% Mismatches: 1
Query Match: 93.78% Indels: 0
DB: 2 Gaps: 0

US-10-625-972-4 (1-2091) x 087518 (1-669)

QY	82	ATGATTCTCTCGGATCCGGCTATGAGAAAGCTGACTAACCGACCCCGCAGTGTTCCT	141
DB	1	MetileValSerAlaSerGlyTyrGluLysLysLeuThrAsnAlaAlaSerValSer	20
QY	142	GTGATTAGCCAGGAGAAATGCGAGTCCAGCAGTACACGATCTGCGCGAGGCTCTGACA	201
DB	21	ValileSerGlnGluGluLeuGlnSerSerGlnTyrHisAspLeuAlaGluAlaLeuArg	40
QY	202	TCAGTAGAGGCTGGATGTTGAAAGTGTACGGGTAAACCGGAGGCTGGAATCAGC	261
DB	41	SerValGluGlyValAspValGluSerGlyThrGlyLysThrGlyLeuGluLeuSer	60
QY	262	ATCCGAGGATGCCAGCCAGTACACGCTGATACGATGATGATGATGATGATGATGATG	321
DB	61	IlEArgGlyMetProAlaSerTyrThrLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu	80
QY	322	AGCAGTGAAGTACTCCCAACGGTTTTCTGCCATGAATACCGGGTTTCATGCCCTCTG	381
DB	81	SerSerAspValThrProAsnGlyPheSerAlaMetAsnThrGlyPheMetProProLeu	100
QY	382	CCCGCCATTGAGGCTTATGAGTTATCAGGGCCGATGTCACACTGTATGCTCTGAT	441
DB	101	AlaAlaIleGluArgIleGluValIleAArgIleProMetSerThrLeuTyrGlySerAsp	120
QY	442	GCATGGCGGCTGTGTGATATATATATATATATATATATATATATATATATATATAT	501
DB	121	AlaMetGlyValValAlaIleIleThrArgLysAsnAlaAspLysTrpLeuSerSer	140
QY	502	GTCAATCCAGGCTGTAATCTGAGGAAGCAAAATGGGGTAAACAGCAGGATTTAAT	561
DB	141	ValAsnAlaGlyLeuAsnLeuGlnGluSerAsnLysTrpGlyAsnSerSerGlnPheAsn	160
QY	562	TTCTGAGCAGTGGTCCCTTGGGATGATTTCTGTCAGCCTCAGGTACCGCGTAGCACA	621
DB	161	PheTrpSerSerGlyProLeuValAspAspSerValSerLeuGlnValArgGlySerThr	180
QY	622	CAACAGCTCAGGTTTCATCGGTACATCACTGAGCGATACAGCAGCGACCGGTATTCCT	681
DB	181	GlnGlnArgGlnGlySerValThrSerLeuSerAspThrAlaAlaThrArgIlePro	200
QY	682	TATCCACCGAGTACAGATTTATATCTTGGTGCAGCTTGTGACTGGAAGCGCTCGGAG	741
DB	201	TyrProThrGlnSerGlnAsnTyrAsnLeuGlyLysLysLysLysLysLysLysLysLys	220
QY	742	CAGGATGTCTCTGGTTGATATGATACCCCGCAGCGTTATGATAACCGGATGGG	801
DB	221	GlnAspValLeuTrpPheAspMetAspThrThrArgGlnArgTyrAspAsnArgAspGly	240

QY	802	CAACTGGGAGTCTGACGGGGGATATGACCGGACCCCTGGCTGATGAGCAAAACAAATT	861
DB	241	GlnLeuGlySerLeuThrGlyTyrAspArgThrLeuArgTyrGluArgAsnLysIle	260
QY	862	TCAGCTGGCTATGATCATCTTTACCTTCGGAACATCGAATCTGTAACCTGGAAC	921
DB	261	SerAlaGlyTyrAspHisThrPheThrPheGlyThrTrpLysSerTyrLeuAsnTrpAsn	280
QY	922	GAGACAGAAAATAAAGGTCGTGAGCTGTACGAGTGTACTGAAAGCGCAGAAATGGGG	981
DB	281	GluThrGluAsnLysGlyArgGluLeuValArgSerValLeuLysArgAspLysTrpGly	300
QY	982	CTTGCCGCTCAGCCCGGAGCTTAAGGAATCGAACCTTATCTCGAATTCATCTACTCT	1041
DB	301	LeuAlaGlyGlnProArgGluLeuLysGluSerAsnLeuLeuLeuLeuLeuLeuLeu	320
QY	1042	ACCCCTCTGGGAGAAATCTCATCTCTGTTACGTTGGGGGCGAGTTTCAGAGCTCGTCCATG	1101
DB	321	ThrProLeuGlyGluSerHisLeuValThrValGlyGlyGluPheGlnSerSerMet	340
QY	1102	AAAGACGGAGTGTCTTGGCCAGCACAGGTCAAACTTTCCGGCAGAAAAGCTGTGCGGTA	1161
DB	341	LysAspGlyValValLeuAlaSerThrGlyGluThrPheArgGlnLysSerTrpSerVal	360
QY	1162	TTTCTGAGGATGAGTGGCATCTCACGGATGTCACCTTCGCTGCTGCGGCGACCGCTAT	1221
DB	361	PheAlaGluAspGluTrpHisLeuThrAspAlaLeuAlaLeuThrAlaGlySerArgTyr	380
QY	1222	GAAATCATGACGAATTCGGGGGACACTTCAGTCCGCGTGCATATCTGTCTGGGATGTG	1281
DB	381	GluHisGlnGluGlnPheGlyHisPheSerProArgAlaTyrLeuValTrpAspVal	400
QY	1282	GCAGATCCCTGGAGCTGAAAGCGGTGTGACCGGGATATAGGACCCAGAAATGGGG	1341
DB	401	AlaAspAlaTrpThrLeuLysGlyValThrThrGlyTyrLysAlaProArgMetGly	420
QY	1342	CAGCTACATAAAGGATTAGTGGTGTCTCCGGCAGGAAAAAACAATCTACTTGGTAAC	1401
DB	421	GlnLeuHisLysGlyLysGlyValSerGlyGlnGlyLysThrAsnLeuLeuGlyAsn	440
QY	1402	CCCGACCTGAAGCGGAGAGAGCGTCAGTTATGAGCTGGGTGTATTACGATAACCC	1461
DB	441	ProAsnLeuLysProGluGluSerValSerTyrGluAlaGlyValTyrTyrAspAsnPro	460
QY	1462	CCCGCTGTAATGCAATGTACAGGTTTTATGACTGACTTCTCCACAGATTTCTCT	1521
DB	461	AlaGlyLeuAsnAlaAsnValThrGlyPheMetThrAspPheSerAsnLysIleValSer	480
QY	1522	TATTCCATAAATGATAACCAATAGCTATGTAACAGCGGAAAGCCCGGTTGCACGGT	1581
DB	481	TyrSerIleAsnAspAsnThrAsnSerTyrValAsnSerGlyLysAlaArgLeuHisGly	500
QY	1582	GTGAAATTCGGCGCACATTCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	1641
DB	501	ValGluPheAlaGlyThrLeuProLeuTrpSerGluAspValThrLeuSerLeuAsnTyr	520
QY	1642	ACCTGGACCCGAAATGAAACCAACCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	1701
DB	521	ThrTrpThrArgSerGluGlnArgAspGlyAspAsnLysGlyAlaProLeuSerTyrThr	540
QY	1702	CTGTAAACACATGTTGTAATGCGAACTGAACTGCGAGATTCACCGAAGAGTGGCATCGG	1761
DB	541	ProGluHisMetValAsnAlaLysLeuAsnTrpGlnIleThrGluGluValAlaSerTrp	560
QY	1762	CTGGTCCCGCTTATCGCGGAAAAACACACGTTTTTCCACCGAATTTATTCGTCACTGAGC	1821
DB	561	LeuGlyAlaArgTyrArgGlyLysThrProArgPheThrGlnAsnTyrSerSerLeuSer	580
QY	1822	GCTGTACAGAAAGATGTATGATGAAAGAGAGATACTGAAACCGCTGGACGGTGGTG	1881
DB	581	AlaValGlnLysLysValTyrAspGluLysGlyGluTyrLeuLysAlaTrpThrValVal	600

Db 454 ProArgAlaTyrLeuValTyrAspAlaGlnAspAsnTrpThrValLysGlyGlyValSer 473
 QY 1315 ACAGGATATAAGGACCCAGAGTGGCGAGCTACATAAGAGGATAGTGTGTCCGG 1374
 Db 474 ThrGlyTyrLysThrProThrLeuAlaGlnLeuHisAsnGlyIleSerGlyValThrGly 493
 QY 1375 CAGGAGAAACAAATCTACTTGTGTAACCCGACCTGAAGCCGGAAGAGACGTCAGTTAT 1434
 Db 494 GlnGlyThrIleThrThrIleGlyAsnProLysLeuGluProGluSerSerValAsnThr 513
 QY 1435 GAGCTGGGTGATTAACAGTAACCCCGCGTCTGAATGCCAATGTACAGGTTTATG 1494
 Db 514 GluValGlyValTyrGlyAsnGluThrGlyPheGlyAlaAsnValThrLeuPheHis 533
 QY 1495 ACTGACTTCTCCAAACAGATTGCTCTATTCCATAAATGATAACACCAATACCTATGTA 1554
 Db 534 AsnArgPheArgAsnLysIleAsnSerValSerIleAspAsnThrThrSerThrThr 553
 QY 1555 AACAGCGAAAGCCCGGTTCACGCGTGTGCAATTTCCGCGACATTCGCGTGTGTCA 1614
 Db 554 AsnValGlyLysAlaIleThrGlnGlyIleGluValAlaSerThrIleProLeuTyrP 573
 QY 1615 GAGGATCTCAGCTGCTCAGTGAATACACCTGACCGAGTGAACAACTGATGTAT 1674
 Db 574 AspAspTrpMetLeuGlyIleAsnThrThrPheThrAspSerGluGlnLysAspGlyAsn 593
 QY 1675 AACAAAGTGGCGGCTGAGTTATACCCCTGAACACATGTCGTAATCGAACTGAACTGG 1734
 Db 594 AsnLysGlyAlaArgLeuThrAsnThrProLysAsnMetValAsnAlaArgLeuAsnTrp 613
 QY 1735 CAGTACCAAGAGTGGCATGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1794
 Db 614 AsnIleAsnGluGlnLeuSerThrTrpLeuLysAlaGluThrArgSerLysThrAlaArg 633
 QY 1795 TTCACCCAGAAATATTCTGCTGAGCGCTGACAGAGAGTGTATGATGAGAAAGGA 1854
 Db 634 PheThrGlnAsnTyrAlaAsnLeuSerAlaAsnLysValValTyrAsnAsnLeuGly 653
 QY 1855 GAATACCTGAAGCTGGAGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1914
 Db 654 SerGluPheLysProPheSerValLeuAsnLeuGlyValAlaTyrLysValThrLysAsp 673
 QY 1915 CTGACGCTGAATCTGCGGTGAATACCTGCTCAACAGAGATTACAGTACGCTGAGCCTG 1974
 Db 674 ValThrLeuAsnGlyAlaValAsnAsnLeuLysAspPheThrArgThrHisIle 693
 QY 1975 TACAGTCCCGTGAAGTAGTACGCTGTATGCGGTGATTACTTCCAGACGGGATCATCAACA 2034
 Db 694 PheAlaValGlyAsnGlyThrThrThrAlaGlyAspPheThrSerSerGlnSerThr 713
 QY 2035 ACAGGATATGTACTGACGCAATTTACTGATGTCGCTGACCTGATCAGTTTC 2088
 Db 714 AlaGlyTyrValValProGlyArgAsnTyrTrpValSerValAsnValAsnPhe 731

RESULT 6

Q8VR68 PRELIMINARY; PRT; 390 AA.
 AC Q8VR68; 2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE R4-like protein.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CFT073;
 RX MEDLINE=21458820; PubMed=11574920;
 RA Rasko D.A., Phillips J.A., Li X., Mobley H.L.T.;
 RT Identification of DNA sequences from a second pathogenicity island of
 RT uropathogenic Escherichia coli CFT073: a probes specific for

RT uropathogenic populations.";
 RL J. Infect. Dis. 184:1041-1049 (2001).
 CC -!- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).
 DR EMBL; AF447814; AAL67380.1; -.
 DR GO; GO:0019867; C:outer membrane; IEA.
 DR GO; GO:0004872; P:receptor activity; IEA.
 DR GO; GO:0005215; P:transporter activity; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR000531; TonB_boxC.
 DR Pfam; PF00593; TonB_dep_Rec; 1.
 KW Membrane; Outer membrane; Receptor; TonB box.
 SQ SEQUENCE 390 AA; 42718 MW; 0B72B0B02B65BA12 CRC64;
 Alignment Scores:
 Pred. No.: 6,16e-104 Length: 390
 Score: 1759.00 Matches: 343
 Percent Similarity: 99.71% Conservative: 0
 Best Local Similarity: 99.71% Mismatches: 1
 Query Match: 47.13% Indels: 0
 DB: 2 Gaps: 0
 US-10-625-972-4 (1-2091) x Q8VR68 (1-390)
 QY 1 ATGGGAATAACCACTCTGGCTTCCTAGTCAATCCCTGCTCGGATTTTCAGCCAGCAGC 60
 Db 1 MetArgIleThrThrLeuAlaSerValValIleProCysLeuGlyPheSerAlaSer 20
 QY 61 ATAGCTGCTGCAGAGGATGTGATGATTTCTCGGCATCCGCTATGAGAAAAGCTGACT 120
 Db 21 IleAlaAlaAlaGluAspValMetIleValSerAlaSerGlyTyrGluLysLysLeuThr 40
 QY 121 AACGACGCGCAGTGTTCCTGATTTAGCCAGAGGAAATGCGAGTCCACCCAGTACCAC 180
 Db 41 AsnAlaAlaAlaSerValIleSerGlnGluLeuGlnSerSerGlnTyrHis 60
 QY 181 GATCTGCGGAGGCTCTGAGATCAGTAGAGGCTGTGATGCTTGAAGTGTACGGGTAAA 240
 Db 61 AspLeuAlaGluAlaLeuArgSerValGluGlyValAspValGluSerGlyThrGlyLys 80
 QY 241 ACCGAGGCGTGGAAATCAGCATCCGAGGAATGCCAGTATACAGCTGATACGATT 300
 Db 81 ThrGlyGlyLeuGluIleSerIleArgGlyMetProAlaSerTyrThrLeuIleLeu 100
 QY 301 GATGCTTCTGTCAGCGGGAAGCAGTACGTCACCTCCCAACGGTTTTCGCGCATGAT 360
 Db 101 AspGlyValArgGlnGlySerSerAspValThrProAsnGlyPheSerAlaMetAsn 120
 QY 361 ACCGGGTTTCATGCCCTCTGCGCCCATTTGAGCGTATTGAGGTATTACGGGGCGCATG 420
 Db 121 ThrGlyPheMetProProLeuAlaAlaIleGluArgIleGluValIleArgGlyProMet 140
 QY 421 TCACACTGTATGGCTCTGATGGAGTGGCGGTGTGTGTAATATCATATTACCAAGAAGAT 480
 Db 141 SerThrLeuTyrGlySerAspAlaMetGlyGlyValValAsnIleIleThrArgLysAsn 160
 QY 481 GCAGACAAATGGCTCTCTCCGTCATTCAGGCGTGAATCTGCAGGAACCAACAATGG 540
 Db 161 AlaAspLysTrpLeuSerSerValAsnAlaGlyLeuAsnLeuGlnGluSerAsnLysTrp 180
 QY 541 GGTAAACAGCAGCGATTAAATTTCTGGAGCAGTGGTGGTCCCTTGTGATGATTCTGTGAGC 600
 Db 181 GlyAsnSerSerGlnPheAsnPheTrpSerSerGlyProLeuValAspSerValSer 200
 QY 601 CTGCAGTACGCGGTAGCACACACGCGTTCATCGGTTCATCATCATCATCATCATCATCAT 660
 Db 201 LeuGlnValArgGlySerThrGlnGlnArgGlnGlySerSerValThrSerLeuSerAsp 220
 QY 661 ACAGCAGCAGCGGTATTTCCTTATCCACAGGAGTTCACAGAAATTAATCTTGTGTGACGT 720
 Db 221 ThrAlaAlaThrArgIleProTyrProThrGluSerGlnAsnTyrAsnLeuGlyAlaArg 240
 QY 721 CTGACTGGAGGCTCGGAGCGAGTGTCTCTGTTGATGATGATGATGATGATGATGATGATG 780

241 LeuAspTyrLysAlaSerGluGlnAspValLeuTyrPheAspMetAspThrThrArgGln 260
 781 CQTATGATACCGGATGGCACTGGGGAGTCTGACGGGGGATATGACGGACCGCTG 840
 261 ArgTyrAspAsnArgAspGlyGlnLeuGlySerLeuThrGlyGlyTyrAspArgThrLeu 280
 841 CQATGATGAGCAAAACAAATTTTCAGCTGGCTATGATCATCTTTCACCTTCGGAACATGG 900
 281 ArgTyrGluArgAsnLysIleSerAlaGlyTyrAspHisThrPheThrPheGlyThrTyr 300
 901 AAATCGTATCTGAATCGAAGCAGACAGAAATAAAGTCGTGCTGTGACGAGTGTA 960
 301 LysSerTyrLeuAsnTyrAsnGluThrGluAsnLysGlyArgGluLeuValArgSerVal 320
 961 CTGAGCGCGCAATGGGGCTTCGGCTGAGCGCGGAGCTTAAGAAATCGAACCTT 1020
 321 LeuLysArgAspLysTyrGlyLeuAlaGlyGlnProArgGluLeuLysGluSerAsnLeu 340
 1021 ATCCTGAATTC 1032
 341 IleLeuAsnSer 344
 |||||
 RESULT 7
 P70847
 ID P70847 PRELIMINARY; PRT; 754 AA.
 AC P70847;
 DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
 DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Exogenous ferric siderophore receptor.
 GN BFA OR B4761.
 OS Bordetella bronchiseptica (Alcaligenes bronchiseptica).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Alcaligenaceae; Bordetella.
 OX NCBI_TaxID=518;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RB50 / ATCC BAA-588;
 RX MEDLINE=22827954; PubMed=12910271;
 RA Parthill J., Sebahia M., Preston A., Murphy L.D., Thomson N.,
 Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
 Cardeno-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
 Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
 Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
 Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
 Leather S., Moulé S., Norbertczak H., O'Neil S., Ormond D., Price C.,
 Rabinowitz E., Rutter S., Sanders M., Saunders D., Seeger K.,
 Sharp S., Simmonds M., Skelton J., Squares R., Stevens K.,
 Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
 RA "An iron-regulated outer-membrane protein specific to Bordetella
 bronchiseptica and homologous to ferric siderophore receptors."
 RL Microbiology 143:135-145(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RB50 / ATCC BAA-588;
 RX MEDLINE=22827954; PubMed=12910271;
 RA Parthill J., Sebahia M., Preston A., Murphy L.D., Thomson N.,
 Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
 Cardeno-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
 Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
 Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
 Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
 Leather S., Moulé S., Norbertczak H., O'Neil S., Ormond D., Price C.,
 Rabinowitz E., Rutter S., Sanders M., Saunders D., Seeger K.,
 Sharp S., Simmonds M., Skelton J., Squares R., Stevens K.,
 Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
 RA "Comparative analysis of the genome sequences of Bordetella pertussis,
 Bordetella parapertussis and Bordetella bronchiseptica."
 RL Nat. Genet. 35:32-40(2003).
 CC -!- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).
 DR EMBL; U56084; AAB51774.1; -;
 DR EMBL; BX640451; CAE35124.1; -;
 DR GO; GO:0019867; C:outer membrane; IEA.
 DR GO; GO:0004872; F:receptor activity; IEA.
 DR GO; GO:0005215; F:transporter activity; IEA.
 DR GO; GO:0006910; P:transport; IEA.
 DR InterPro; IPR01064; Crystallin.
 DR InterPro; IPR000531; TonB box.
 DR Pfam; PF00593; TonB_dep_Rec; 1.
 DR PROSITE; PS00225; CRYSTALLIN BETAGAMMA; 1.
 DR PROSITE; PS00430; TONB DEPENDENT REC 1;
 KW Membrane; Outer membrane; Receptor; TonB box; Complete proteome.

SQ SEQUENCE 754 AA; 82628 MW; A2929E37EE58AE61 CRC64;
 Alignment Scores:
 Pred. No.: 9,39e-68 Length: 754
 Score: 1192.00 Matches: 275
 Percent Similarity: 52.41% Conservative: 127
 Best Local Similarity: 35.85% Mismatches: 243
 Query Match: 31.94% Indels: 122
 DB: 16 Gaps: 19
 US-10-625-972-4 (1-2091) x P70847 (1-754)
 Qy 22 TCGCTAGTCAATCCCTGCTCCGATTTTTCAGCCAGCAGCATAGCTGCTGCAGAG----- 75
 Db 32 AlaLeuAlaIleAlaLeuGlyAlaAlaGlnThrAlaSerProGlnGluAsp 51
 Qy 76 -----GATGTGATGATGCTCGGCATCCGGCTATGAGAAAAAG 114
 Db 52 ThrLeuGlyIleThrGlnMetAspThrValValThrAlaSerGlyPheGluGlnGlu 71
 Qy 115 CTGACTAACCGCAGCCGCTCTGATGATTTCTGTGATTTAGCCAGGAGAAATGCGATCCAGCCAG 174
 Db 72 IleLysAsnAlaProLysIleSerValIleThrArgGluGlnLeuGluSerLysPro 91
 Qy 175 TACACGATCTGGCGGAGGCTCTGATCAGTACGTAGAGGGTGTGATGTTGAACTGGTACG 234
 Db 92 PheHisAsnLeuAlaAspAlaValAlaAspValGluGlyValSerValGluArgGly--- 110
 Qy 235 GGTAAACCGGAGGCTCGAAATCAGCATCCGAGGAATGCCAGCAGTTACAGCTGATA 294
 Db 111 GlyLysAlaGlyMetAsnIleSerIleArgGlyLeuProSerAspTyrThrLeuVal 130
 Qy 295 CTGATTCATGCTGTGCTGTGCTGATGCGATGCGGCGGTGTGGTGAATATCATTTACC 471
 Db 131 LeuValAspGlyLysArgLeuSerGlnAsnSerSerGlyAlaArgProAsnGlyPheGly 150
 Qy 352 GCCATGAATACCGGTTTCATGCCCCCTCTCGCGCCATTCAGCGTATTGAGGTATTACAGG 411
 Db 151 AspValAspThrAsnProMetSerAlaIleAspArgIleGluValValArg 170
 Qy 412 GGGCGGATGTCACACATGCTGATGCGATGCGGCGGTGTGGTGAATATCATTTACC 471
 Db 171 GlyProMetSerThrLeuTyrGlySerAspAlaMetGlyGlyValIleAsnIleThr 190
 Qy 472 AGAAGAAATCAGACAAATGGCTCTCTTCGTCATTCAGCGGCTGATCTGCAGGAAAGC 531
 Db 191 ArgLysValAlaArgGluTyrThrGlyGlnValThrLeuAspGlyThrAlaGlnGlyAsp 210
 Qy 532 AACAAATGGGTACACAGCAGCCAGTTTAATTTCTGGAGCAGTGTCCCTTTGCGATGAT 591
 Db 211 AsnArgTyrGlyAsnAsnTyrGlySerSerPheTyrLeuSerGlyProLeuGlnThrAsp 230
 Qy 592 TCTCTCAGCTGCGAGGTACGGGTAGCACACACAGCGT-----CAGGGTTCATCG 642
 Db 231 LysLeuGlyLeuSerLeuArgGlyGlyLeuTyrArgArgLeuSerAlaHisGlySer--- 249
 Qy 643 GTCACATCATGCGCATACAGCAGCGCGTATTCTTATCCCGAGTACACAGAT 702
 Db 250 -----TyrProAlaAsnGlnAlaGlu 256
 Qy 703 TAT----- 705
 Db 257 TyrAspSerGlyAspTyrSerGlyAspIleAlaSerPheSerGlyLeuGlyAspSerLeu 276
 Qy 706 -----RATCTGGTGGCAGCTCTTTCATGCGAGCGCTCGGAGCAGGAGTGTCTCGTTT 759
 Db 277 GlnArgAsnValGlyLeuArgLeuAlaLeuThrProAsnArgAsnHisAspIleLeuPhe 296
 Qy 760 CATATGATACCAACCCCGCAGCTTATCATTAACCGGATGGCAACTGGGAGTCTGACG 819
 Db 297 AspValAspAlaAsnTyrGlnThrPheAspAsnAlaAsnGlyGluLeuGlyThrLeuAsn 316
 Qy 820 -----GGGGGATATGACCGGACCCCTCGCGCTATGACGCGA 852

Db 317 AlaAspValAlaProAsnArgGlnGlyGlyTyrGluProGluMetLysPheAsnArg 336
 QY 853 AACAAATTCA-----GCTGGCTATGATCATACTTTCACCTTCGGAACATCG 300
 Db 337 GlnArgTyrAlaLeuThrHisLeuGlyArgTyrAspGlyAlaIleSerSerAspThr--- 355
 QY 901 AATCGTATCTGAACGACGACGACGACGACGACGACGACGACGACGACGACGACG 960
 Db 356 -----SerLeuLeuTyrAspThrThrGluThrIleGlyArgThrAsnProMetSerThr 373
 QY 961 CTGAAGCGCGACAAATGGGGCTTGCCTGACGCGCGAGCTTAAGAAATCGAACCTT 1020
 Db 374 ProArgGlnProSer-----AspGlyGluLysArgGluLeuGluTyrGluAsnTrp 390
 QY 1021 ATCTGAAATTCATCTGCTTACCCCTCTG---GGAGAAATCTCATCTGTTACGGTGGG 1077
 Db 391 ValPheAspThrLysTrpThrMetProLeuPheAsnAspArgHisAsnLeuThrMetGly 410
 QY 1078 GCGGAGTTTCAGAGCTCGTCCATGAAGACGAGTGTCTCTCCAGCACAGCTGAAACT 1137
 Db 411 GlyGlnTrpArgGluGlnLysPheLysAsp-----ThrLeuValSerAlaProLeuAsn 428
 QY 1138 TTCGGGAGAAAGCTGGTGGTATTGTCAGGATGAGTGGCATCTCCAGATGCACTT 1197
 Db 429 LeuArgGlnTyrGlnTrpAlaLeuPheAlaGluAspGluTrpArgIleValAspAspLeu 448
 QY 1198 GCGTGAAGTGGGCGGCGCTATGACATCATGAGCAATTCGGGGACACATTCAGTCGG 1257
 Db 449 AlaLeuThrMetGlyAlaArgTyrAspArgAsnGluGlnPheGlyLysTrpSerPro 468
 QY 1258 CGTGCATATCTGCTGGGATGCGGAGATGCTCGACGCTGAAAGCGGCTGTGACACG 1317
 Db 469 ArgGlyTyrLeuValTrpAsnAlaThrProAlaTrpThrValLysGlyLysValSerLys 488
 QY 1318 GGATATAGCCACCCAGATGGGCGAGTACATAAAGGATAGTGGTGTCCGGGAG 1377
 Db 489 GlyTyrLysThrProAspIleAsnLeuMetThrAspGlyIleLeuGlyLysAlaGln 508
 QY 1378 GGAAGAAACAAATCTACTTGTAAACCCGACCTGAAAGCGGAGAGAGCTCATGATGAG 1437
 Db 509 GlyThrMetProLeuLeuGlyAsnSerGlnLeuLysProGluSerSerThrSerGlu 528
 QY 1438 GCTGGGGTGTATACGATACCCCGGCTGCTGATGCAATGCCAATGTCACAGTTTATGACT 1497
 Db 529 LeuGlyValLeuPheAspAspGlyGlyGlyLeuThrGlyAsnLeuThrGlyPheHisThr 548
 QY 1498 GACTTCTCCAAACAGATGCTCTTATCCATAAATGAT----- 1536
 Db 549 LysPheLysAspLysIleAspThrGlnAsnValProAsnCysLeuAlaAlaGlyGlyPro 568
 QY 1537 -----AACACC 1542
 Db 569 ValProGlyCysLeuAspLeuGlyValTrpGluArgAsnGlyValProValAlaAsnPhe 588
 QY 1543 AATAGCTATGTAACAGCGGAAGCGCGTTCACGCTGCGAATTTCCGCGCACATG 1602
 Db 589 SerGlnArgValAsnValAspThrAlaThrIleGlnGlyPheGluLeuGlyGlyArgIle 608
 QY 1603 CGCTGTGTGTCAGAGATGTCAGCTGTCTCACTGAATTCACCTGGACCGGAGTGAACAA 1662
 Db 609 ProLeuPhe---GluGlyTrpSerPheSerGlyAsnTyrThrLeuThrAlaSerGluIle 627
 QY 1663 CGTGATGTCGATACAAAGTGGCGGCTGAGTATACCCCTGAACACATGTCGTAATCGG 1722
 Db 628 ThrSerGlyAlaLysGlnGlyGlnProLeuGlySerGlnProArgHisSerLeuAsnLeu 647
 QY 1723 AAATGAACTGGCAGATCAACCGAAGAGTGGCATCATGCTGGTGGCTGATTCGCGGG 1782
 Db 648 GlyLeuAsnTrpArgValAsnGluArgPheAsnAlaTrpValArgGlyGluTyrArgAla 667
 QY 1783 AAACACACAGTTTACCCAGAAATTTTCGTCACTGAGCGGTGTACAGAAAGATGTAT 1842

Db 668 Lys-----GlnPheAsnAspMetAsnTrpGluLysGluGlnValPhe 681
 QY 1843 GATGAGAAAGGAGAAATACCTGAAAGCCTGGACGGTGGTGGATGAGTCTGTCGTGGAAG 1902
 Db 682 -----TyrSerProTyrTrp---LeuAlaSerLeuGlyGlySerTyrVal 695
 QY 1903 ATGACGGATGCCCTGACGCTGAATGCTGCGTGAATACCTGCTCAACAAGGATTACAGT 1962
 Db 696 LeuAsnLysAsnValThrLeuSerAlaSerValTyrAsnLeuPheAspLysAsnPheVal 715
 QY 1963 GACGTGACCTGTGACGTCGCGGTAAAGAGT-----ACGCTGTATGCCCGT 2007
 Db 716 AspTyrGlyProThrLysValGlyThrSerAlaProThrAlaAlaThrSerTyrSerAsn 735
 QY 2008 GATTACTTCCAGCGGATCATCAACACAGGATATGTGATCTGACCTGAGCGAAATTTACTGG 2067
 Db 736 SerTyrArgGln-----ValLeuGluGlyArgArgLeuTrp 747
 QY 2068 ATGTCGCTGAACTATCATGTTTC 2088
 Db 748 ValSerAlaAsnIleThrPhe 754
 RESULT 8
 Q9PPG3 PRELIMINARY; PRT; 696 AA.
 AC Q9PPG3; 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Putative iron uptake protein.
 GN CFRA OR C00755.
 OS Campylobacter jejuni.
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
 OC Campylobacteraceae; Campylobacter.
 OX NCBI_TaxID=197;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCCT 11168;
 RX MEDLINE=20150912; PubMed=10688204;
 RA Parkhill J., Wren B.W., Mungall K., Kettle J.M., Churcher C.,
 RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
 RA Jagels K., Karlyshev A.V., Moule S., Palien M.J., Penn C.W.,
 RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
 RA Whitehead S., Barrall B.G.;
 RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
 RT reveals hypervariable sequences."
 RL Nature 403:665-668(2000).
 DR EMBL; AL139076; CAB73021.1; .
 DR PIR; D81346; D81346.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0004872; F:receptor activity; IEA.
 DR GO; GO:0005215; F:transporter activity; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR000531; TonB_boxC.
 DR Pfam; PF00593; TonB_dep_Rac; 1.
 KW Complete proteome.
 SQ SEQUENCE 696 AA; 77554 MW; 31E09CB15187388C CRC64;
 Alignment Scores:
 Pred. No.: 1,79e-62 Length: 696
 Score: 1109.00 Matches: 255
 Percent Similarity: 54.80% Conservative: 150
 Best Local Similarity: 34.51% Mismatches: 248
 Query Match: 29.72% Indels: 86
 DB: 16 Gaps: 16
 US-10-625-972-4 (1-2091) x Q9PPG3 (1-696)
 QY 1 ATCGGATAACCACTCTGGCTCCGTAGTATCCCTGCTCGGATTTTCGACGACGACG 60
 Db 1 MetLysLysIleCysLeuSerValCysAlaIleGlyLeuLeuSerSerAsnAlaIleSer 20
 QY 61 ATAGCTGCTCAGAGGATGTGATGTCTCGGCATCCGGCTATGAGAAAAGTGACT 120

Db 21 GlnAsnValGluLeuAspSerSerIleValSerAlaSerGlyPheThrGlnAspIleLys 40
 Qy 121 AACGACGCGCCAGTGTCTCTGATTAGCCAGGAGGATTCAGCTCCAGCCAGTACCAC 180
 Db 41 GluAlaProAlaThrIleasnValIleThrLysLysGluLeuGlnSerLysProTyrArg 60
 Qy 181 GATCTGGGGAGGCTCTGAGATCAGTAGAGGTGTGGATGTGGAAAGTGGTACCGGTAA 240
 Db 61 AspValAlaGluAlaAlaAspIleProGlyValAspLeuTyrAlaSerLysGlyLys 80
 Qy 241 ACCGAGGCGTGAATCAGCATCCGAGGAATGCCAGGAGTACACCTCTGATCTGATT 300
 Db 81 ThrGlySerTyrAsnIleThrMetArgGlyIle---ThrGlyTyrThrLeuValLeuIle 99
 Qy 301 GATGTGTCTGTCAGGGCGGAAGAGTGAAGTACTCCCAACGGTTTCTGCCCATGAAT 360
 Db 100 AspGlyArgGlnGlyIleGlyGluValGlyProAsnGlyPheAsnGluIleSer 119
 Qy 361 ACCGGGTTCATCCCTCTGCGCCATTCAGCGCTATTGAGGTATTACAGGGCGCGATG 420
 Db 120 AsnSerPheLeuProIleSerSerIleGluArgIleGluValIleLysGlyProMet 139
 Qy 421 TCCACACTGTATGGTCTGATCCGATGGCGGTGTGGTGAATATCATTTACCAGAAAGAT 480
 Db 140 SerThrLeuTyrGlySerGluAlaLeuGlyGlyValValAsnIleIleThrLysLysVal 159
 Qy 481 GCAGCAATATGCTCTCTCCCTCAATGCAGGGCTGAATCTCCAGGAAGCAAAATGG 540
 Db 160 SerAspLysTrpGluThrSerValSerLeuAspAlaLeuLeuAsnGluAsnLysAspTrp 179
 Qy 541 GGTAAACAGCAGCAGTAAATTTCTGGAGCAGTGTCCCTTGTGGATGATCTCTCAGC 600
 Db 180 GlyAsnThrTyrGlyThrSerIleTyrSerSerGlyProLeuMetAsnAspLysLeuGly 199
 Qy 601 CTCAGGTCAGCGGTAGCACAAACAGCGTCAAGGTTCATCGGTACATCACTGAGCGAT 660
 Db 200 LeuThrLeuArgPheArgGluPheTyrArgGlnGlnSerAsnVal---GluPheThrAsn 218
 Qy 661 ACAGCAGGACCGCTATT-----CCTTATCCACGAGGTACAGAAATAT 705
 Db 219 GlySerGlyArgValGlnGlyAspGlnAlaGlnSerProThrLysAlaAsnAsnPhe 238
 Qy 706 AATCTTGGTCAGCTTCTGACTGGAAGCGTCGAGCAGGATGTCTCTGGTTGATATG 765
 Db 239 AsnIleGlyThrArgIleSerTyrLeuAlaAsnAspTyrAsnThrPheIlePheAspIle 258
 Qy 766 GATACACCGCGGAGCGTATGATAACCGGGATGGCACTG-----807
 Db 259 AspPheSerArgAsnHisTyrAspAsnLysGlnGlyLeuGlyThrIleThrSerPro 278
 Qy 808 -----GGGAGTCTGACGGGGGATATGACCGGACCTCGCTATGAGCGAAAC 855
 Db 279 GlyArgThrProGlySerLeuThrGlyGlyTyrAlaAspIleMetGluValAspLysPhe 298
 Qy 856 AAAATTTCAGTGGCTATGATCATCTTTCACCTTCGGAACATGGAATTCGTATCTGAA 915
 Db 299 ValThrTyrLeuSerHisGluGlyValTyrGluAsnPheSerIleThrSerGlyLeuGln 318
 Qy 916 TGAACGACGACAGAAATAAGTCTGAGCTGTACGAGTGTACGAGTGTACTGAGCGCACAA 975
 Db 319 TyrAsnArgValSerAsnAspGlyArgGluValVal-----330
 Qy 976 TGGGGCTTCCCGCTCAGCGC-----CGGAGCTTAAAGAAATCGAACCTT 1020
 Db 331 ---GlyGlnSerThrGlnProPheLeuGlyGluAsnArgAspIleValAlaGluAspIle 349
 Qy 1021 ATCTGAATTCATCTGTTACCTCTGGGAGATCTCATCTGTTACGTGGGGGCG 1080
 Db 350 IleLeuAspThrLysSerValIleProLeuGlyGlnSerHisIleLeuSerValGlyGly 369
 Qy 1081 GAGTTTCAGAGCTCGCTCCATGAAGACCGAGTGTCTCTGCCAGCAGCTGAAATTC 1140

Db 370 GluTyrArgLeuGluLysMetGlnAspLysIle-----AlaSerProThrAsnPhe 386
 Qy 1141 CGCAGAGAAAGCTGTCGGTATTGTCGAGGATGAGTGGCATCTCAGGATGCACTTGGC 1200
 Db 387 AspGlnTyrLeuLeuAlaIlePheAlaGluAspGluTyrSerIleLysAspLeuArg 406
 Qy 1201 CTGACTCGCGGACGCGCTATGAACATCATGACCAATTCGGGGACACATTCAGTCCGCGT 1260
 Db 407 LeuThrPheGlyAlaArgTyrAsnHisGluIlePheGlyAsnAsnValSerProArg 426
 Qy 1261 GCATATCTGTCTGGGATGTGGCAGATGCTGGACGCTGAAGCGCGTGTGACACGGGA 1320
 Db 427 AlaTyrValValTyrAsnProThrAsnGluLeuThrLeuLysGlyValSerThrGly 446
 Qy 1321 TATAAGCACCCAGATGGGCGAGCTACATAAAGGATAGTGTGTCTCCGGCAGGGA 1380
 Db 447 PheArgThrProTyrAlaAsnArgLeuIleAsnGlyThrTyrSerTyrSerGlyGlnGly 466
 Qy 1381 AAAACAAATCTACTGTGTAACCCCGACCTGAACCGGAGAGAGCGTCAGTTATGAGGCT 1440
 Db 467 ArgPheProThrTyrGlyAsnProAspLeuLysGluGluThrSerLeuAsnTyrGluIle 486
 Qy 1441 GGGGTGTATTACGATAACCCCGCGGTCTGAATGCCAATGTCACAGGTTTATGACTGAC 1500
 Db 487 AlaAlaIleTyrAsnAsnAspLeu---PheTyrValSerAlaThrGlyPheLeuThrAsn 505
 Qy 1501 TTCTCCCAACAGATTGTCTTCTTATTCATAAAT-----1533
 Db 506 PheLysAspLysIleSerSerGlnSerTyrAsnAsnSerGluProIleProGlyIleGly 525
 Qy 1534 -----GATAACACCAATAGCTATGTAACAGCGGAAAGCCCGGTGCACGGT 1581
 Db 526 ThrCysAspAlaAspArgCysSerArgAlaIleAsnHisGlyLysValGluTyrLysGly 545
 Qy 1582 GTGGAATTT---GCCGACATTCGCGCTGTGTGTCAGAGGATGTACGCTGTCACTGAAT 1638
 Db 546 ValGluLeuGlyAlaGlyIleSerProLeu-----AspAsnLeuAsnValAsnPheAla 563
 Qy 1639 TACACCTGGACCCGAGTGAACACCGTATGGTGATTAACAAA-----GGTGGCCCG 1689
 Db 564 TyrThrTyrLeuAspThrGluValLysGluAlaGlnAspArgSerValIleGlyLysPro 583
 Qy 1690 CTGATGTTATACCTCTGAACACATGGTGAATCCGAACTGAACCTGGCAGATCACCGAAG 1749
 Db 584 GluGlnAspSerLeuLysHisAsnIleMetLeuLysThrGluTyrSerPheTyrAsnLys 603
 Qy 1750 GTGGCATCATGCTGGTGGTGCC-----CGTTATCGCGGAAAAACACCCAGT 1794
 Db 604 IleThrProThrIleLysGlyGluTrpGlnIleAspArgTyrMetGlyAspThr-----621
 Qy 1795 TTCACCCAGAAATTATTCGTCACTGAGCGCTGTACAGAAAGAGTGTATGATGAGAAAGGA 1854
 Db 622 -----AsnIleAsnArg 625
 Qy 1855 GAATACCTGAAAGCCTGGACGGTGTGGATGTCAGGTCTGTGCGAAGATGACGGATGCC 1914
 Db 626 GluTyrTyrLysAspIlePheLeuAlaSerMetGlyValArgTyrAspIleAsnLysGln 645
 Qy 1915 CTGACGCTGATGCTGGCGTGAATTAACCTCTCAACAGGATTAACAGTAC---GTGAGC 1971
 Db 646 TrpSerIleAsnAlaIleTyrAsnLeuPheAspLysSerPheThrAsnGlyTrpGlu 665
 Qy 1972 CTGTACAGTCCGCTAGAGATGCTGTATGCCGCTGATCTTCCAGACGGGATCATCA 2031
 Db 666 SerTyrAlaSerGlySerGlySerThrTrpValAsnThrTyr-----679
 Qy 2032 ACAACAGGATATGTATGATACCTGAGCGAAATTTACTGGATGTGCGTGAACACTACGTT 2088
 Db 680 -----AsnArgIleGluGluGlyArgArgMetTyrIleSerIleAsnGlyAsnPhe 696
 RESULT 9
 Q7VI87
 ID Q7VI87 PRELIMINARY; PRT; 728 AA.

AC Q7V187;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein cfrA.
 GN CfrA OR H0721.
 OS Helicobacter hepaticus.
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
 OC Helicobacteraceae; Helicobacter.
 OX NSB1_TaxID=32025;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 51449 / 3B1;
 RC MEDLINE=22709201; PubMed=12810954;
 RA Suerbaum S., Josenhans C., Sierzenbach T., Drescher B., Brandt P.,
 RA Bell M., Droege M., Partmann B., Fischer H.-P., Ge Z., Hoerster A.,
 RA Holland R., Klein K., Koenig J., Macko L., Mendz G.L., Nyakatura G.,
 RA Schauer D.B., Shen Z., Weber J., Frosch M., Fox J.G.;
 RT "The complete genome sequence of the carcinogenic bacterium
 RT Helicobacter hepaticus.";
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7901-7906 (2003).
 DR EXBL; AE017146; AAP77318.1; -;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 728 AA; 79692 MW; 95683019216E9718 CRC64;

Alignment Scores:
 Pred. No.: 3 5e-62 Length: 728
 Score: 1104.50 Matches: 257
 Percent Similarity: 54.58% Conservative: 136
 Best Local Similarity: 35.69% Mismatches: 252
 Query Match: 29.60% Indels: 75
 DB: 16 Gaps: 16

US-10-625-972-4 (1-2091) x Q7V187 (1-728)

QY 55 AGCAGATACCTGCTCAGAGGATGTG-----ATGATTCTCGCATCC 99
 DB 42 SerSerSerSerAlaThrArgAsnValLysLeuGlnLysSerIleValThrAlaAlaSer 61
 QY 100 GGCTATGAGAAAAGCTGACTAACGCGCGCCAGTGTCTGTGATTAGCCAGGAGAA 159
 DB 62 GlySerGluLysAsnValIleAspAlaProAlaSerValAsnIleIleThrLysGluGlu 81
 QY 160 TTCAGTCCAGCAGTACACAGTCTCGCGGAGCTCTGAGTACAGTACAGGCTGTGGAT 219
 DB 82 LeuGluGlnLysProTyrArgAspLeuGlyGluAlaLeuLysGluValProGlyValSer 101
 QY 220 GTTGAAGTGGTACGGTAAACCGGAGGCTGGAATCAGCATCCGGAATGCCAGCC 279
 DB 102 LeuAspGluThrSerAsnLysLeuGlyAlaSerAlaIleSerIleArgGlyMetProSer 121
 QY 280 AGTTACAGCTGATCTGATGTGTTGTTGTCAGCGCGGAGAGTACAGTGCCTG---ACT 336
 DB 122 GlyTyrThrLeuPheLeuValAspGlyLeuArgGlnAsnProSerGlyAspValAlaThr 141
 QY 337 CCCAAC---GGTTTCTGCGCATGATACCGGTTCTATCCCTCTGATCCCTGCGCCCATTTAG 393
 DB 142 AlaAsnLeuGlyAlaGlyValTyrAsnThr---PheIleProMetGlyAlaIleGlu 160
 QY 394 CGPATTGAGTTATCAGGGGCGGATGTCCACATCTATGGTCTGTATGCGATGGCGGT 453
 DB 161 ArgIleGluValIleLysGlyProMetSerThrLeuTyrGlySerAspAlaIleGlyGly 180
 QY 454 GTCGTGAATATCATCCAGGAAGATCGACAAATGGCTCTCTCCGTCAATCGAGG 513
 DB 181 ValValAsnIleIleThrLysProIleThrAspLysIleThrPheSerPheGlnSerGln 200
 QY 514 CTGAATCTCAGGAAACCAAAATGGGTAAACAGCAGCAGCTTTAAATTTCTGGAGCAGT 573
 DB 201 IleIleValProGlnSerThrPheGlyAsnThrTyrGlnAsnSerLeuIleThrGln 220
 QY 574 GGTCCCTGTGTGATGAT---TCTGTCAGCTCAGGTACGGTAGCAGCAACAGCGT 630
 DB 571 GlnGlyValGluSerSerPheGlyIleLysProIleTyrGly-----ValGlyPheAsp 598

DB 221 GlyGlyLeuThrLysHisLeuGlyLeuThrPheArgValArgGlnIleHisLysAsnPro 240
 QY 631 CAGGTTTCATCGTCATCATCTACTAGCGATACAGCAGCAGCGGTATCTTATCCACG 690
 DB 241 SerAspArgProLysAsnAspLeuGlyGlnAspValSerThrPhe-----PheGlyThr 258
 QY 691 GAGTCACAGATTATAATCTTGGTGCACTCTTGGTACCTGGAAGCGCTCGGAGCAGGATGTG 750
 DB 259 LysPheSerLeuLeuAsnValGlyThrArgLeuThrIleProAspSerLysAsnMet 278
 QY 751 CTCTGGTTTGAATGATATACCCCGCAGCGGTATGATTAACCGGATGGCACTGGGG 810
 DB 279 LeuTyrAlaAspValAspTyrThrSerAsnTyrAspAsnArgLysGlyGluIleGly 298
 QY 811 AGTCTG-----ACGGGGGATATACCGGACCCCTGCGCTATGAGCGA 852
 DB 299 ThrLeuGlyValAspValAsnGlyArgGlyGlyTyrGluLysAsnValGlyValAspLys 318
 QY 853 AACAAATTTTCAGTGTATGATCATATCTTCCACCTTCGGAACATGGAATCGTATCTG 912
 DB 319 IleLeuGlyAlaIleGlyHisLysGlySerTyrAspPheGlyThrTyrLysAsnThrLeu 338
 QY 913 AACTGGACGAGACAGAAATAAGTCTGAGCTTCTAGCGAGTGTACTGAGAGCGGAC 972
 DB 339 GlnPheMetArgThrAsnAsnThrGlyArgLeuValAlaGlyAsnThrAsnSerProAsn 358
 QY 973 AAATGGGGGCTTGGCGGTACGCGCGGAGCTTAAGGAATCGAACCTTATCTCGAATTC 1032
 DB 359 -----IleGlyLysAsnArgAspAlaSerAsnAspValIleAlaAspThr 374
 QY 1033 TTACTGTTACCTCTGGGAGATCTCATCTGTTACGTTGGTGGGGCGGAGTTTCAGAGC 1092
 DB 375 ArgLeuLeuLeuProLeuGlyGluSerAsnIleLeuAsnValGlyAlaGluTyrArgPhe 394
 QY 1093 TCGTCCATGAAAGCAGGAGTTGCTCTTCCAGCAGCAGGTGAAACTTTCGCGCAGAAAAGC 1152
 DB 395 GluAsnTyrTyrAsp-----LeuAlaAlaThrProAlaSerHisAsnArgHisThr 411
 QY 1153 TGTGCGGATTTGCTGAGATGATGTCATCTCAGCGATGCACTTGGCGTGCATCGGGGC 1212
 DB 412 PheAlaLeuPheAlaGluAspGluTyrAsnIlePheGluProLeuThrPheThrLeuGly 431
 QY 1213 AGCGCTTATGAACATCATGACGATTCGGGGACACTTCAGTCCGGTGGTGCATATCTGGTC 1272
 DB 432 AlaArgTyrAsnHisAsnAspLysPheGlySerAsnIleSerProArgAlaIleVal 451
 QY 1273 TGGATGTGCGAGATGCTCGACGCTGAAAGCGGTGTGACACCGGATATAAGGCACCC 1332
 DB 452 TyrGluIleLeuAspGlyPheAlaLeuLysGlyValAlaThrGlyTyrLysAlaPro 471
 QY 1333 AGANTGGGCGAGTACATAAAGGATGTGTTGTCGGGCGAGGAAAAACAATCTA 1392
 DB 472 TyrAlaAsnGlnLeuIleAspAlaValTyrGlyTyrGlySerGlnGlyThrLeuAlaPhe 491
 QY 1393 CTTGCTAACCCACCTGAAAGCGGAGAGAGCGTCTGATGAGTGGTGGGTGTATTAT 1452
 DB 492 LeuGlyAsnProAsnLeuLysProGluThrSerValSerTyrGluIleGlyThrIlePhe 511
 QY 1453 GATTAACCCCGCGGTGTAATGCAATGTCACAGGTTTATGACTGCTCTCCACAAG 1512
 DB 512 AspGly---ArgTyrIleAspPheSerLeuMetPhePheArgSerAsnPheLysAspLys 530
 QY 1513 ATTGCTCTTATTCATAAATGATAACACCAATAGC----- 1548
 DB 531 IleGluAsnThrArgValAsnLysAsnAlaAsnGlyAlaPheTyrGluIleThrCysGln 550
 QY 1549 -----TATGTAAACAGCGGAAAGCCCGGTG 1575
 DB 551 AsnTyrGlyThrAsnThrGlnCysAsnLeuAlaIleAsnAlaAspSerAlaPheSer 570
 QY 1576 CAGGCTGTGGAATTTGCC---GGCATTCTCGCTGTGGTTCAGAGGATGTTCACGCTGTCA 1632
 DB 571 GlnGlyValGluSerSerPheGlyIleLysProIleTyrGly-----ValGlyPheAsp 598

1021 ATCTGTAATTCATTACTGCTTACCCCTCTCGGAGAAATCTCATCTGTTACGGTGGGGGG 1080
 350 ILeuAspThrLysSerValIleProLeuGlyGlnSerHisIleLeuSerValArgGly 369
 1081 GAGTTTCAGAGCTCGTCCATCAAGACGAGGTGTCCTCCAGCAGCAGTGAATCTTC 1140
 370 GlutylargLeuGlyMetGlnAspLysIle-----AlaSerProThrAsnPhe 386
 1141 CGGAGAGAAAGCTGGTTCGATTTGCTGAGCATGAGTGGCATCTCCAGGATGACATTGCG 1200
 387 AspGlnTyrLeuLeuAlaIlePheAlaGluaspGluTyrSerIleLysAspLeuArg 406
 1201 CTGACTCGGCGAGCCGCTATGAAATCATCATGAGCAATTCGGGGACATTCAGTCGCGT 1260
 407 LeuThrPheGlyAlaArgTyrAsnHisHisGluIlePheGlyAsnValSerProArg 426
 1261 GCATATCTGCTGGGATGGCGAGATGCTCGACGCTGAAGCGGTGTGACACCGGA 1320
 427 AlaTyrValValTyrAsnProThrAsnGluLeuThrLeuLysGlyValSerThrGly 446
 1321 TATAAGCACCCAGATGGGCGAGCTACATAAAGGATAGTGGTGTCTCCGGCGAGGA 1380
 447 PheArgThrProTyrAlaAsnArgLeuIleAsnGlyThrTyrSerTyrSerGlyGlnGly 466
 1381 AAAACAAATCTACTGTGTAACCCCGACCTGAACCGGAGAGAGCTCAGTTATGAGGCT 1440
 467 ArgPheProThrTyrGlyAsnProAspLeuLysGluGluThrSerLeuAsnTyrGluIle 486
 1441 GGGGTGTATTACGATAACCCCGCGCTGTAATGCCAATGTACAGTGTATTTATGACTAC 1500
 487 AlaAlaIleTyrAsnAsnAspLeu---PheTyrValSerAlaThrGlyPheLeuThrAsn 505
 1501 TTCTCCACAGAGTTGCTCTTATTCATAAAT----- 1533
 506 PheLysAspLysIleSerGlnSerTyrAsnSerGluProIleProGlyIleGly 525
 1534 -----GATAACCACTAGCTATGTAACAGCGGAGAGCCCGGTGCACCGT 1581
 526 ThrCysAspAlaAspArgCysSerArgAlaIleAsnHisGlyLysValGluTyrLysGly 545
 1582 GTGGAATT---CCGGGCAATTCGCGTGTGTGTCAGAGATGTACGCTGTCATCTGAAT 1638
 546 ValGluLeuGlyAlaGlyIleSerProLeu-----AspAsnLeuAsnValAsnPheAla 563
 1639 TACACTGGAACCGCAAGTGAACCACTGATGTGTAACAAA-----GGTGGCGG 1689
 564 TyrThrTyrLeuAspThrGluValLysGluAlaGlnAspArgSerValIleGlyLysPro 583
 1690 CTGAGTTATACCCCTGAACACATGGTGAATGCGAACTGAATCGGAGATCACCGAAG 1749
 584 GluGlnAspSerLeuLysHisAsnIleMetLeuLysThrGluTyrSerPheTyrAsnLys 603
 1750 GTGGCATCATGCTGGTGGTCC-----CGTTATCGGGGAACACACCGT 1794
 604 PheThrProThrIleLysGlyGluTrpGlnIleAspArgTyrMetGlyAspThr----- 621
 1795 TTCACCCAGAAATTATTCGTCACCTGAGCGCTGTACAGAGAAAGTGTATCATGAGAAAGGA 1854
 622 -----AsnIleAsnArg 625
 1855 GAATACCTGAAGCTCGCGTGAATACCTGCTCAACAGGATTACAGTGAC---GTGAGC 1914
 626 GluTyrTyrLysAspIlePheLeuAlaSerMetGlyValArgTyrAspIleAsnLysGln 645
 1915 CTGACCTGAATGCTCGCGTGAATACCTGCTCAACAGGATTACAGTGAC---GTGAGC 1971
 646 TrpSerIleSerAlaAlaIleTyrAsnLeuPheAspAsnSerPheThrAsnGlyTrpGlu 665
 1972 CTGTACAGTCGGGTGAAGATGACGTGTATGCGGTGATTAATCTCCAGCAGGATCATCA 2031
 666 SerTyrAlaSerGlySerGlySerThrTrpValAsnThrTyr----- 679

2032 ACAACAGCATATGTATCATCTACCTAGCGAGAAATTAAGTGTGCTGTAATCATCTTC 2088
 680 -----AsnArgIleGluGluGlyArgArgMetTyrIleSerIleAsnGlyAsnPhe 696
 RESULT 11
 Q8ZG14 PRELIMINARY; PRT; 665 AA.
 ID Q8ZG14;
 AC Q8ZG14;
 DT 01-MAR-2002 (TRENBLrel. 20, Created)
 DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
 DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
 DE Putative outer membrane protein (Putative outer membrane
 iron/siderophore receptor).
 GN YPO1313 OR YIUR OR Y2872.
 OS Yersinia pestis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Yersinia.
 OX NCBI_TaxID=632;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CO-92 / Biovar Orientalis;
 RX MEDLINE=21470413; PubMed=11586360;
 RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
 RA Prentice M.B., Sebaihia M., James K.D., Churcher C., Mungall K.L.,
 RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
 RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
 RA Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
 RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.,
 RT "Genome sequence of Yersinia pestis, the causative agent of plague";
 RL Nature 413:523-527(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KIM5 / Biovar Mediaevalis;
 RX MEDLINE=22137863; PubMed=12142430;
 RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
 RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
 RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
 RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
 RA Perry R.D.;
 RT "Genome sequence of Yersinia pestis KIM.";
 RL J. Bacteriol. 184:4601-4611(2002).
 DR EMBL; AJ414147; CAC90143.1; -;
 DR EMBL; AE013889; AAM86423.1; -;
 DR PIR; AD0160; AD0160.
 DR GO; GO:0016020; C.membrane; IEA.
 DR GO; GO:0004872; P.receptor activity; IEA.
 DR GO; GO:0005215; P.transporter activity; IEA.
 DR GO; GO:0006810; P.transpor; IEA.
 DR InterPro; IPR00531; TonB_boxC.
 DR Pfam; PF00593; TonB_dep Rec; 1.
 DR PROSITE; PS00430; TONB DEPENDENT REC 1; 1.
 KW Hypothetical protein; Receptor; Complete proteome.
 SQ SEQUENCE 665 AA; 73909 MW; F5ACA23F30B06B71 CRC64;

Alignment Scores:
 Pred. No.: 2,58e-57 Length: 665
 Score: 1028.00 Matches: 251
 Percent Similarity: 52.12% Conservative: 130
 Best Local Similarity: 34.34% Mismatches: 224
 Query Match: 27.55% Indels: 126
 DB: 16 Gaps: 20

US-10-625-972-4 (1-2091) x Q8ZG14 (1-665)

QY 22 TCGGTAGTATCCCTCTCGGATTTTCAGCC-----AGCAGCATAGCTGCTGCAGAG 75
 Db 19 AlaAlaValIleSerSerGlnGlyTyrAlaAlaGluLysThrAsnThrAlaThrProThr 38
 QY 76 GATGTGATGATGTTCTCGGATCCGCTATGAGAAAGCTGACTAACGACGCGCAGT 135
 Db 39 AspThrMetValValThrAlaSerGlyPheGlnArgIleGlnAspSerAlaAlaSer 58

01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 Iron-regulated outer membrane virulence protein homolog.
 VP2602.
 DB Iron-regulated outer membrane virulence protein homolog.
 GN VP2602.
 OS Bacterium paraheamolyticus.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrio.
 OC NCBI_TaxID=670;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RM2210633 / Serotype O3:K6;
 RX MEDLINE=22508454; PubMed=12620739;
 RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
 RA Iijima Y., Najima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
 RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.,
 RT "Genome sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism
 distinct from that of *V. cholerae*."
 RL Lancet 361:743-749 (2003).
 DR EMBL; AF005082; BAC60865.1;
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0004872; F:receptor activity; IEA.
 DR GO; GO:0005215; F:transporter activity; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR00531; TonB_boxC.
 DR Pfam; PF00593; TonB_dep_Rec; 1.
 DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
 DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
 KW Complete proteome.
 SQ SEQUENCE 552 AA; 71947 MW; D05C2AA617508C9E CRC64;

Alignment Scores:
 Pred. No.: 7,71e-57 Length: 652
 Score: 1020.50 Matches: 255
 Percent Similarity: 52.57% Conservative: 123
 Best Local Similarity: 35.47% Mismatches: 232
 Query Match: 27.34% Indels: 109
 DB: 16 Gaps: 24

US-10-625-972-4 (1-2091) x Q87LK9 (1-652)

QY 16 CTGGCTTCCTAGTATCTCCCTCTCGGATTTTCAGCCAGCAGCATAGCTGCTGCAGAG 75
 DB 15 LeuAlaPheValSerValProHisVal---PheAlaAsnAspSerValSerLysValGlu 33
 QY 76 GATGTGATGATTTCTCGGATCCGCTATGAGAAAGAAAGTCACTAAGCCAGCCGCGAGT 135
 DB 34 ThrVal---ValValThrAlaSerSerTyrgluGlnSerGlnAlaAspAlaProAlaSer 52
 QY 136 GTTTCGTGATTAAGCCAGGAGGATTCAGTCCAGCCAGTACACGATCTGCGGAGGCT 195
 DB 53 IleSerValIleSerArgGluGluLeuLeuAspSerArgTyrgluValThrAspAla 72
 QY 196 CTGAGATCAGTAGAGGCTGTGATGTGAAAGTGTACGGTAAACCGGAGGCTGGAA 255
 DB 73 LeuLysSerValProGlyValValValThrGlyGlyGlyAspThrThr-----Asp 89
 QY 256 ATCAGATCCGAGGATGCGCAGGATTCACGCTGATCTAGTGTGATGCTGCTGCTGCTG 315
 DB 90 IleSerIleArgGlyMetGlySerLysTyrgluValIleLeuValAspGlyLysArgGln 109
 QY 316 GCGGGAAGCAGTGCATGCTCCCAACGGTCTTCTCT---GCCATGAATACCGGTTTCATG 372
 DB 110 ---SerThrArgGluThrArgProAsnSerAspGlyProGlyIleGluGlnGlyTrpLeu 128
 QY 373 CCCCCTCTGGCCGATGAGGCTATGAGGTTATCAGGGGCGGATGTCACATGCTGAT 432
 DB 129 ProProLeuGlnAlaIleGluArgIleGluValIleArgGlyProMetSerThrLeuTyrg 148
 QY 433 GGTCTGTGATGCGGCGCTGTGGGTAATATCAATACAGAAAGAAATGCAGACAAATGG 492
 DB 149 GlySerAspAlaIleGlyValIleAsnValIleThrArgLysAspAlaLeuGluTrp 168
 QY 493 CTCTCTTCGTCATGCGGCTGATCTGCGAGGAAGCAAAATGGGGTAAACGAGCAGC 552

DB 169 ThrGlyAsnValGlnLeuGlyThrValIleGlnIleAsnSerArgSerGlyGlyGln 188
 QY 553 CAGTTTAAATTTCTGGAGCAGTGGTCCCTTGTGGATGATTCTGTCAACCTCGAGGTACGC 612
 DB 189 SerAlaAsnPheValAsnGlyProLeuAla---GluAsnLeuLeuGlnValTyrg 207
 QY 613 GGTAGCACAAACAGCGTTCATCGGTTCATCGGTTCATCGGTTCATCGGTTCATCGGTTC 672
 DB 208 GlyGlnTyrgluAlaArgGluGluAspAsp----- 217
 QY 673 CGTATTCCTTATCCACCGGAGTCAAGAAT---TATAATCTTGTGTGACGCTCTTGTACTGG 729
 DB 218 ---IleAspTyrglyTyrgluAspLysAspMetGlnSerIleSerLysLeuIleTyrg 236
 QY 730 AAGCGCTCGGAGCAGGATGCTCTGTTTGAATATGATATGATATGATATGATATGATATG 789
 DB 237 GlnIleAsnAspArgHisSerValGlnLeuGluGlyThrSerAlaGln----- 253
 QY 790 AACCGGATGGCAACTG-----GGGAGTCTGACGGGGGATAT 828
 DB 254 SerArgGlyAsnValGlyLeuSerValProThrThrGlyCysArgArgGlyCysGlu 273
 QY 829 GACCGACCTCGCTATGACGCGAAACAAATTTACGCTGGCTATGATATATCTTCAACC 888
 DB 274 AspSerLeuAsnGluTyrgArgAsnTyrgValThr-----LeuSerHisThr----- 289
 QY 889 TTCGAAACATCGAAA-----TCGTATCTGAACTCGAAGCAGACGAGACAGAA 930
 DB 290 ---GlyGluTrpGluLeuLeuGlyAsnSerAspThrTyrgluGlnArgGluGluSerGlu 308
 QY 931 AATAAAGCTCGTGTGAGCTTGTACGCGAGTGTACTGAAGCGCAGCAAAATGGGGCTTGGCGGT 990
 DB 309 AsnLysSerArgGluMetThr----- 315
 QY 991 CAGCGCGGAGCTTAAGGAATCGAACCTTATCTGTGAATTCATCTACTGTTACCCCTCTG 1050
 DB 316 -----IleValAsnThrThrPheLysSerSerLeuValLysGlyLeu 329
 QY 1051 GGAGAAATCTCATCTGTACGGTGGGGGCGAGTTTCAGAGCTCGTCCATGAAGACGGA 1110
 DB 330 Gly---AlaHisThrLeuThrThrGlyValAspAlaThrHisAlaGluLeuGluAspPhe 348
 QY 1111 GTTCTCCTCCGACGACAGGTGAACCTTCCCGCAGAAAGC---TGGTCGGTATTGCT 1167
 DB 349 ThrSerAsnLysAlaSerLysThrLysAlaSerAsnThrGlnTrpAlaValPheIle 368
 QY 1168 GAGATCAGTGGCATCTCACGGATGCACTTGCCTGCTGCTGCGGCGAGCCGCTATGAACAT 1227
 DB 369 GluAspGluTrpLysIleAlaGluProPheSerLeuThrLeuGlyArgLeuAspHis 388
 QY 1228 CATGAGCAATTCGGGGGACACTTCAGTCCGCTGCATATCTGCTCTGGATGTGGCAGAT 1287
 DB 389 AspGluAsnTyrglyAlaHisPheSerProArgValTyrglyValTrpArgValAlaPro 408
 QY 1288 GCTGGACGCTGAAGCGGTGTGACCGGATATAAGCCAGCAGCAGATGGGGCAGCTA 1347
 DB 409 AlaTrpThrValLysGlyGlyValAlaThrGlyPheArgSerProGlnLeuArgGluIle 428
 QY 1348 CATAAAGGGATTAAGTGTGTCTCCGGCAGGGAACAAATCTACTTGTGAACCCCGAC 1407
 DB 429 ThrProGlyTrpAlaGlnValSerGlyGlyGly-----AsnIleTyrglyAsnProAsp 446
 QY 1408 CTGAGCCGGAAGAGCGCTCAGTTATGAGGTGGGTGTATACATTAACCCCGCGGT 1467
 DB 447 LeuAspProGluThrSerLeuAsnLysGluIleSerValLeuTyrgluGlnGlyAspSerGly 466
 QY 1468 CTGAAGGCCAATCTCAGAGTTTATGACTCTCTCAACAAAGATTCTC----- 1518
 DB 467 LeuAspValThrLeuThrAlaPheHisAsnGluPheLysAspLysIleThrArgValVal 486
 QY 1519 -----TCTTATTCATAATGATAACACCAATAGCTAT----- 1551
 DB 487 CysProAspThrIleCysThrAspGlyProAsnGlnTrpGlyAlaAspProThrTyrg 506

```

QY 1552 GTAAACCGGAAAGCCCGGTTGCAAGCTGGGAATTTGCCGACATTCGCGCTGG 1611
Db 507 ILeuValAspGluAlaValThrGlnGlyValGluAlaThrLeuAlaLysProLeu--- 525
QY 1612 TCAGAGGATGTCACGGCTGCTCACTGAATACACCTGGACCCGAGTGAACACGTTGAT 1671
Db 526 ThrGluThrIleTyrLeuSerSerTyrThrPheThrAspSerGluGlnLysThrGly 545
QY 1672 GATAACAAGGTGGCCCGCTAGTATACCCCTGAACACATGGTGAATGCCAACTGAAC 1731
Db 546 GluTyrLysGlyMetProLeuGlnLysProLysHisLeuPheAsnValAspValThr 565
QY 1732 TGGCAGATCACCGAAGAGTGCATCATGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1791
Db 566 TrpGlnThrThrAspAsnLeuSerTrpThrLysValThrTyrArgGlyLysGluMet 585
QY 1792 CGTTTCACCCAGAAATATTGCTACATGAGCGCTGTACAGAAAGTGTATGATGAGAAA 1851
Db 586 AspProValThrGlyProSerArgAsnSerIleValGlu----- 598
QY 1852 CGAAGATACCTGAAGCCTGGACGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1911
Db 599 -----ProAlaThrThrPheValAspAlaGlyValThrTyrGlnLeuThrAsp 614
QY 1912 GCCCTGACGCTGAATGCTGCGGTGAATTAACCTGCTCAACAAGGATTACAGTACGCTGAGC 1971
Db 615 AsnThrLysIleLysGlyAlaIleTyrAsnLeuPheAspGluAspIleAsn----- 631
QY 1972 CTGTACAGTCCGGTAAAGATGACGCTGTATGCCGTGATTACTTCCAGCGGATCATCA 2031
Db 632 ---TyrLys-----GluTyr----- 635
QY 2032 ACAACAGGATATGTATACCTGAGCGAAATTAATCTGCTGCTGCTGCTGCTGCTGCTGCT 2088
Db 636 -----GlyTyrValGluAspGlyArgTyrTrpLeuGlyLeuAspValAlaPhe 652

```

RESULT 13

```

Q93RD3 ID Q93RD3 PRELIMINARY; PRT; 652 AA.
AC STRAIN=WP1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Iron-regulated outer membrane virulence protein homolog.
GN IRGA.
OS Vibrio parahaemolyticus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=670;
RN [1]
RP SEQUENCE FROM N.A.
RC Tanabe T., Funahashi T., Miyoshi S., Yamamoto S.;
RT "Identification and characterization of the Vibrio cholerae irgAB
RT homologs in Vibrio parahaemolyticus".
RL Submitted (DEC-2000) to the EMBL/GenBank/DBSJ databases.
CC -!- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).
DR EMBL; AB032655; BAB60695.1; -.
DR GO; GO:0019867; C:outer membrane; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0008910; P:transport; IEA.
DR InterPro; IPR000531; TonB_box.
DR Pfam; PF00593; TonB_dep_Rec; 1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW Membrane; Outer membrane; Receptor; TonB box.
SQ SEQUENCE 652 AA; 71957 MW; ACE945161D926FD3 CRC64;

```

Alignment Scores:

```

Pred. No.: 1.39e-56 Length: 652
Score: 1016.50 Matches: 255

```

```

Percent Similarity: 52.43% Conservative: 122
Best Local Similarity: 35.47% Mismatches: 233
Query Match: 27.24% Indels: 109
DB: 24 Gaps: 24

```

US-10-625-972-4 (1-2091) x Q93RD3 (1-652)

```

QY 16 CTGGCTTCGGTAGTCAATTCCTGCTCGGATTTTCAGCCAGCAGCATACCTGCTGCAGAG 75
Db 15 LeuAlaPheValSerValProHisVal---PheAlaAsnAspSerValSerLysMetGlu 33
QY 76 GATGTGATGATGCTCTCGCATCCGCTATGAGAAAAGCTGACTAACCCAGCCGCGACT 135
Db 34 ThrVal---ValValThrAlaSerSerTyrGluGlnSerGlnAlaAspAlaProAlaSer 52
QY 136 GTTCTGTGATTAGCCAGGAGGAATTCAGTCCAGCCAGTACCAGCATCTGGCGGAGGCT 195
Db 53 IleSerValIleSerArgGluLeuAspSerArgTyrTyrArgAspValThrAspAla 72
QY 196 CTGAGATCAGTAGAGGCTGGATGTTGAAAGTGGTACCGGTAAACCCGAGGCGCTGGAA 255
Db 73 LeuLysSerValProGlyValValValThrGlyGlyGlyAspThrThr-----Asp 89
QY 256 ATCAGATCCGAGGAATGCCAGCCACTTACACCTCATACTGATTCATGCTGCTCCTCAG 315
Db 90 IleSerIleArgGlyMetGlySerLysTyrThrLeuIleLeuValAspGlyLysArgGln 109
QY 316 GCGGAGAGCAGTCACTGCTCCCAACGGTTTTCT---GCCATGAATACCCGGGTTCATG 372
Db 110 ---SerThrArgGluThrArgProAsnSerAspGlyProGlyIleGluGlnGlyTrpLeu 128
QY 373 CCCCTCTGCGCCCATGAGCGTATTGAGTTATCAGGGGCGCGATGTCACACTGTAT 432
Db 129 ProProLeuGlnAlaIleGluArgIleGluValIleArgGlyProMetSerThrLeuTyr 148
QY 433 GGCTCTGATCGGATGGCGGTGGTGAATATCATTACCAGAAAGAAATGCAGACAAATGG 492
Db 149 GlySerAspAlaIleGlyValIleAsnValIleThrArgLysAspAlaLeuGluTrp 168
QY 493 CTCTCTTCGTCATCGAGGCTGAATCTCCAGAAAGCAACAAATGGGTACACAGCAGC 552
Db 169 ThrGlyAsnValGlnLeuGlyThrValIleGlnGluAsnSerArgSerGlyGlyGluGln 188
QY 553 CAGTTTAATTTCTGGAGCAGTGGTCCCTTGTGGATGATTCCTCAGCCCTGCGAGTACGC 612
Db 189 SerAlaAsnPhePheValAsnAlaProLeuAla---AspAsnLeuLeuGlnValTyr 207
QY 613 GGTAGCACACACAGCGCTCAGGTTTCATCGGTCACTCATCTGAGCCGATACAGCAGGCAG 672
Db 208 GlyGlnTyrThrAlaArgGluGluAspAsp----- 217
QY 673 CGTATTCCTTATCCCGAGGAGTCAAGAAAT---TATAATCTTGGTGCACGCTTGTACTGG 729
Db 218 ---IleAspTyrGlyTyrGluAspLysAspMetGlnSerIleSerSerLysLeuIleTyr 236
QY 730 AAGCGCTCGGAGCAGGATGCTCTGTTTGTATGATATGATACCCACCGGACGCTTATGAT 789
Db 237 GlnIleAsnAspArgHisSerValGlnLeuGluGlyGlyThrSerAlaGln----- 253
QY 790 AACCGGATCGGCNACTG-----GGGAGTCTGACCGGGGGGATAT 828
Db 254 SerArgArgGlyAsnValGlyLeuSerValProThrThrGlyCysArgArgGlyCysGlu 273
QY 829 GACCGGACCTCGCTATGAGCGAAACAAAATTTCACTGGCTATGATCATCTTTTCCACC 888
Db 274 AspSerLeuAsnGluTyrArgArgAsnTyrValThr-----LeuSerHisThr----- 289
QY 889 TTCGAAACATGGAAA-----TCGTATCTGAATCGGAACAGACAGAGAA 930
Db 290 ---GlyGluTrpGluLeuLeuGlyAsnSerAspThrTyrLeuGlnArgGluGluSerGlu 308
QY 931 AATAAAGGTGCTGAGCTTGTCAGCGAGTGCTACTGAAGCGCGCAAAATGGGCGCTTGCCTGG 990

```


QY 19 GCTTCGGTAGTCATTCCCTGCTCGGATTTTCAGCCAGCAGCATAGCT-----GCTGCA 72
Db 26 AlaAlaMetAlaGlyThrAlaLeuGlyPheAlaValProAlaPheAlaGlnAlaSer 45
QY 73 GAG-----GATGTGATGATGCTCGGATCCTCGGCTATGAGAAAGCTG 117
Db 46 GluGlyAsnThrValLeuGlnGlnIleValValThrAlaSerGlyPheGluGlnAsnVal 65
QY 118 ACTAACGAGCGCGGAGTTCGTGATGATTAGCCAGAGGAATTCAGCTCAGCCAGTAC 177
Db 66 LysAspAlaProAlaSerIleThrValThrArgGluAspLeuGluLysGlySerTyr 85
QY 178 CAGCATCTGGCGGAGCTCTGAGATCAGTACAGAGGTGTGGATGTGAAAGTGTACGGGT 237
Db 86 ArgAspLeuThrAspAlaLeuArgGluValGlnGlyValSerVal-----ThrGly 102
QY 238 AAAACCGAGAGGCTGGAATCAGATCCGAGGAATGCCAGCAGTTACACCTCATACTG 297
Db 103 IleAlaAsnGluLysAspValPheIleArgGlyLeuProGlyAlaTyrThrLeuIleLeu 122
QY 298 ATTGATGCTGTTCGTCAGGCGGAGGAGCAGTGCAGTCCCAACGGTTTTCTGCCATG 357
Db 123 ValAspGlyLysArgGln---SerThrArgAspAlaArgThrAsnGlyAsnSerGlyPhe 141
QY 358 AATACCGGGTTCATGCCCTCTCGCCGCTATGAGCGTATGAGCTATGAGTTATCAGGGGCG 417
Db 142 GluGlnSerPheValProValSerAlaIleGluArgIleGluValValArgGlyPro 161
QY 418 ATGTCACACATGATGGCTCTGATCGGATGGCGGTGTGGTGAATCATTCATACAGAAAG 477
Db 162 MetSerSerLeuTyrGlySerAspAlaMetGlyGlyValIleAsnIleIleThrArgLys 181
QY 478 AATCAGACAAATGGCTCTCTTCCTCAATGACGGGCTGAATCTCCAGGAAGACAA 537
Db 182 ValGlyAspValTyrSerGlySerValThrThrGluGlyThrValGlnGlnHisSerLys 201
QY 538 TGGGTAAACAGCAGCGCTTAATTTCTGACAGTGTTCCTCTGTGATGATTCCTGTC 597
Db 202 PheGlyAsnSerGlyGlnValSerTyrPheAlaAsnGlyProIleLeuLysAspGlnLeu 221
QY 598 AGCTTCAGTACGGGTAGCAGACACAAACAGCTCAGGGTTTCATCGTCCATCATCTAGC 657
Db 222 GlyLeuGlnLeuTyrGlyArgGlyPheThrArgGlyGluAspArgIle-----LeuAsn 239
QY 658 GATACAGCAGCAGCGGTATTCCTTATCCACGAGTACAGAAATTAATCTTGGTGA 717
Db 240 GlyThrThrGly-----AlaLysGluTyrAspPheAsnGly 251
QY 718 CGTCTTGACTGGAAGCGTCGAGCAGGATGTGCTCTGTTGTGATATGATACACCGCG 777
Db 252 ArgLeuThrPheThrProAsnGluAspHisAspIleTyrLeuGluGlyLysThrArg 271
QY 778 CAGCGTTATGATAACCGGATGGCACTGGGAGTCTGACGGGGGATATGACCGGACC 837
Db 272 LeuArg-----ArgAspAlaGluProGlyAspThrLeuAlaAlaThrAspAlaAsn 288
QY 838 CTCGGTATGACGGAAC---AAAAATTCAGCTGGCTATGATCATCATCTTCACCTTCGGA 894
Db 289 GlyThrTyrAsnThrAsnThrArgAspHisTyrSerLeuSerHisThrGlyArgTyrGly 308
QY 895 -----ACATGGAAATCGTATCTGAACTGGAAACGAG 924
Db 309 ProThrThrSerLeuPheSerPheGlnGluTyrAlaGluArgThrAsnPheThrArg 328
QY 925 ACAGAAAATAAGTGTGAGCTGTGACGAGTGTACTGAGCGGACAAATGGGGGCTT 984
Db 329 AsnThrArgThrGlyArg-----Val 335
QY 985 GCCGGTCAGCGCGGAGCTTAAGGAATCGAACCTTATCTCTGAATTCATTCTGCTTACC 1044
Db 336 ThrGluAsnProArgSerProGluIleArgAsnThrValLeuAspGlyLysPheThrThr 355
QY 1045 CCT-----CTGGAGAAATCTACTCTGGTTACGGTGGGGGCGAGTTTCAGAGCTCGTCC 1098

Db 356 ProPheGluLeuPheGlyAsnHisThrLeuValThrGlyGlyGlnTyrPheGluAlaArg 375
QY 1099 ATGAAAGACGGAGTTGCTCTCTCCAGCAGACA-----GGTGAACACTTTCGGGAGAAAGC 1152
Db 376 LeuThrAspGlnAsnProGlyArgArgThrAspArgAspGluThrPheSerAlaThrGln 395
QY 1153 TGTCTCGGTATTTGCTGAGGATGAGTGGCATCTCAGGATGCATCTGCGCTGCTGCTGGCGC 1212
Db 396 TrpAlaLeuPheLeuGluAspGluTyrArgIleValAspAsnPheAlaLeuLeuThrGlyGly 415
QY 1213 AGCCGTATGAACATCATGCAATTCGGGGGACACTTCAGTCCGCGTGCATATCTGGTC 1272
Db 416 LeuArgLeuAspAsnHisGlnLysTyrGlyAsnHisPheSerProArgLeuTyrGlyVal 435
QY 1273 TGGGATGTGGCAGATCCCTGGACGTGAAGGCGGTGTGACACCGGATATTAAGGACCC 1332
Db 436 TrpSerAlaThrGluGluLeuThrIleLysGlyValSerThrGlyPheArgAlaPro 455
QY 1333 AGAATGGGCGAGCTACATAAGGGATAGTGTGTCTCGGGCAGGAGAAACAAAT--- 1389
Db 456 GluIleArgGlnIleAlaProGlyTyrAlaTyrThrThrGlyGlyGlyCysSerTyr 475
QY 1390 -----CTACTGTGTAAACCCGACCTGAAGCGCGGAGAGAGC 1425
Db 476 GlyProSerGlyThrCysGlyValIleIleGlyAspProAsnLeuGluAlaGluLysSer 495
QY 1426 GTCAGTTATCAGGCTGGGTGTATTACGATAACCCCGCTGTGATGCAATGTGCACA 1485
Db 496 ThrSerTyrGluIleAlaLeuThrAspAsn---GlyAspIleAlaLeuGlyAlaThr 514
QY 1486 GGTATTATGACTGACTCTTCCAAACAAGATT----- 1515
Db 515 TyrPheTyrThrAspPheLysAspLysIleSerAsnAlaLeuValLeuAsnProAspGly 534
QY 1516 -----GTCCTTTATTCATAAAT 1533
Db 535 ThrProAlaArgTyrSerGluAspArgAsnTyrArgLeuTyrTyrAsnTyrAsnIleAsp 554
QY 1534 GATTAACCAATACCTATGTAACAGCGGAAAGCGCGGTGTACAGGTGTGGAAATTTGCC 1593
Db 555 Asp-----AlaIleIleGlnGlyValGluLeuThr 564
QY 1594 GGCACTATGCGCTGTGG-----TCAGAGATGTACGCTGCATGAAATTCACCTGG 1647
Db 565 AlaThr-----TyrTyrAlaThrSerGluLeuThrLeuArgGlyAsnTyrThrTyr 581
QY 1648 ACCGAAAGTGAACACGTGATGTGTATAACAAAGGTCCGCGCTGAGTATACCCCTGAA 1707
Db 582 ThrHisSerGluGlnLysThrGlyAspTyrGluGlyPheProLeuAlaArgThrProGlu 601
QY 1708 CACATGCTGAATGCAAACTGAATGCGCAGATCACC-----GAAGAGGTGGCATATGG 1761
Db 602 HisMetAlaAsnLeuAspGlyAspTyr---ValThrProIleAspGlyLeuGluAlaTyr 620
QY 1762 CTGGGTGCCGCTTATCCGGGAAAAACA-----CCAGTTTCACCCAGAAATAT 1809
Db 621 ValSerLeuAsnTyrHisGlySerGluIleAsnAlaGlyProArgIleGlyAlaAsnGly 640
QY 1810 TCCTCACTGAGCGCTGTACAGAAAGAAAGTGTATGATGAGAAAGGAGAAATACCTGAAAGCC 1869
Db 641 ThrProValThr-----IleAsnGlyGlnSerGlyArgLysTyrAspAla 655
QY 1870 TGGACGCTGTGTGATGAGGTCTGTCTGTGAAGATGACGATGCCCTGACGCTGAATGCT 1929
Db 656 TyrThrThrLeuAspIleGlyAlaLysTyrAlaValAlaGluAsnValAspLeuAsnAla 675
QY 1930 CGGTGTAATACCTGCTCAACAAGGATTACAGTACGCTGAGCCTGTACGCGCGTAAG 1989
Db 676 AlaValTyrAsnValPheAspLysAsp-----Val 684
QY 1990 AGTACGCTGTATGCGGTGATTACTTTCAGACGGGATCATCAACACAGGATATGTGATA 2049


```

Db      460 GlyAlaSerIleMetTyrGlyAsnArgAspLeuLysProGluThrSerValThrGluGlu 479
QY      1438 CTGGGGTGATATACGATAACCCCGCGCTGTAATGCCAATGTACAGGTTTATGACT 1497
Db      480 IleGlyIleIleTyrSerAsnAspSerGlyPheSerAlaSerAlaThrLeuPheAsnThr 499
QY      1498 GACTTCCTCAACAAGATTGCTCTATTCCATAAATGAT----- 1536
Db      500 AspPheLysAsnLysLeuThrSerTyrAspIleGlyThrLysAspProValThrGlyLeu 519
QY      1537 AACACCAATAGTATGTAACAGCGGAAGCCCGGTGACCGGTGCGAATTTGCCGCG 1596
Db      520 AsnThrPheIleTyrAspAsnValGlyGluAlaAsnIleArgGlyValGluLeuAlaThr 539
QY      1597 ACATTGCCGCTGTGGTCAGAGGATCTCAGCGTGCACCTGAATTACACCTGGACCCGAAGT 1656
Db      540 GlnIleProValTyr---AspLysTrpHisValSerAlaAsnTyrThrPheThrAspSer 558
QY      1657 GAACAACGTGATGTGTATAAC-----AAAGTGGCCGCTGAGT 1695
Db      559 ArgArgLysSerAspAspGluSerLeuAsnGlyLysSerLeuLysGlyGluProLeuGlu 578
QY      1696 TATACCCCTGAACACATCGTGAATCGAACTGAACTGGCAGATCACCGAAGAGGTGCA 1755
Db      579 ArgThrProArgHisAlaAlaAsnAlaLysLeuGluIleTyrAspTyrThrGlnAspIleThr 598
QY      1756 TCATGCTGGTGGTCCCGTTATCGCGG----- 1782
Db      599 PheTyrSerSerLeuAsnTyrThrGlyLysGlnIleTyrAlaAlaGlnArgAsnGlyAla 618
QY      1783 AAACACCAACGTTTCACCCAGAAATTATTGCTCAGTACGCGCTGTACAGAAAGTGTAT 1842
Db      619 LysValProArgValArgAsnGlyPheThrSerMet----- 630
QY      1843 GATGAGAAGGAGATACCTGAAGCCTGGACGGTGGTGGATGTCAGGTCTGTCGTGGAAG 1902
Db      631 -----AspIleGlyLeuAsnTyrGln 637
QY      1903 ATGACGGATGCCCTGACGCTGAATGCTGCGGTGAATAACCTGCTCAACAAGGATTACAGT 1962
Db      638 IleLeuProAspThrLeuIleAsnPheAlaValLeuAsnValThrAspArgLysSerGlu 657
QY      1963 GACGTGAGCCTGTACAGTCCCGTAAGAGTACGCTGTATGCCGGTGATTACTTCCAGACG 2022
Db      658 AspIleAspThrIleAsp-----GlyAsnTrp----- 666
QY      2023 GGATCATCAACAACAGGATATGTGATACCTGAGCGAAATTAATGATGTCGCTGAACATAT 2082
Db      667 -----GlnValAspGluGlyArgArgTyrTyrAlaAsnValArgVal 680
QY      2083 CAGTTC 2088
Db      681 SerPhe 682

```

Search completed: October 13, 2004, 11:54:20
Job time : 175 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: October 13, 2004, 11:42:27 ; Search time 42 Seconds

(without alignments)
9577.928 Million cell updates/sec

Title: US-10-625-972-4

Perfect score: 3732

Sequence: 1 atggaataaacactctggc.....cgctgaactatcagttctga 2091

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 566732

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -MODEL=frame+...-DEV=xlh

-Q=cn2_1/USPTO spoil/US10625972/runat 12102004_130106_26831/app_query.fasta_1.2247
-DB=PIR_78 -OPMT=fastan -SUFFIX=trp -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DICALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10625972 @CN_1_1_54 @runat 12102004_130106_26831 -NCPU=6 -ICPU=3
-NO MAP -LARGEBLOCK=100 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR_78.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3632	97.3	696	2 H90798	Iha adhesin [impor
2	3632	97.3	696	2 H85607	probable receptor
3	1109	29.7	696	2 D81346	probable iron upta
4	1052.5	28.2	652	2 D82317	iron-regulated out
5	1052.5	28.2	652	2 S25265	outer membrane pro
6	1028	27.5	665	2 AD0160	probable outer mem
7	993	26.6	702	2 AF3038	exogenous ferric s
8	993	26.6	707	2 E98247	outer membrane pro
9	841.5	22.5	663	2 AG0782	colicin I receptor
10	836.5	22.4	659	2 G91009	colicin I receptor
11	832.5	22.3	663	1 QREGIC	colicin I receptor
12	827.5	22.2	659	2 A85954	hypothetical prote
13	738	19.8	653	2 F83405	probable TonB-depe
14	629	16.9	746	2 G90706	hypothetical prote

15	629	16.9	746	2 B85557	hypothetical prote
16	625	16.7	746	1 QRCFC	ferrienterochelin
17	604.5	16.2	742	2 AC0574	ferrienterobactin
18	603	16.2	742	2 H83529	siderophore recept
19	592.5	15.9	746	2 A40636	ferric enterobacti
20	580.5	15.6	726	2 AG0837	TonB-dependent out
21	541.5	14.5	614	2 A98241	hypothetical prote
22	541.5	14.5	614	2 F86088	hypothetical prote
23	528.5	14.2	614	1 QRCBT	vitamin B12 recept
24	519	13.9	625	2 AD0476	probable vitamin B
25	510.5	13.7	614	2 AB0935	vitamin B12 recept
26	509.5	13.7	611	2 A82358	TonB-dependent rec
27	477	12.8	733	2 A87521	TonB-dependent rec
28	448.5	12.0	650	2 B87466	metal chelate oute
29	440	11.8	599	2 AC3334	hypothetical prote
30	407	10.9	638	2 AG2219	hypothetical prote
31	396.5	10.6	883	2 C83385	probable TonB-depe
32	395	10.6	616	2 B83485	heme utilization/c
33	391	10.5	660	2 D91176	outer membrane hem
34	391	10.5	660	2 B86022	probable TonB-depe
35	373.5	10.0	700	2 F64897	hypothetical prote
36	371.5	10.0	700	2 H85732	heme transport pro
37	371	9.9	698	2 B82443	hypothetical prote
38	369.5	9.9	700	2 G90885	hypothetical prote
39	364.5	9.8	715	2 H90377	hypothetical prote
40	364.5	9.8	719	2 F85824	probable outer mem
41	359.5	9.6	764	2 AC2075	outer membrane hem
42	356	9.5	802	2 H30555	probable ferrichro
43	352.5	9.4	713	2 E91118	probable iron comp
44	352.5	9.4	713	2 D85363	hemin receptor pre
45	350.5	9.4	686	2 S28042	

ALIGNMENTS

RESULT 1

H90798
Iha adhesin [imported] - Escherichia coli (strain O157:H7, substrain RIMD 0509952)
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C:Accession: H90798
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: H90798
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-696 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA834783.1; PID:G13360820; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: ECs1360
C:Superfamily: ferrienterochelin receptor; TonB-dependent receptor amino-terminal homolo

Alignment Scores:	Pred. No.:	Length:	Matches:
Score:	7.89e-242	3632.00	695
Percent Similarity:	99.86%	Conservative:	0
Best Local Similarity:	99.86%	Mismatches:	1
Query Match:	97.32%	Indels:	0
DB:	2	Gaps:	0

US-10-625-972-4 (1-2091) x H90798 (1-696)

QY	1	ATCGAATAACCACTCTCGCTTCGCTAGCTATTCCTCTCGGATTTCAGCCAGCAGC	60
Db	1	MetargilleThrLeuAlaSerValIleProCysLeuGlyPheSerAlaSer	20
QY	61	ATAGCTGCTGAGAGATGATGATTCCTCGGCATCCGCTATGAGAAAAGCTGACT	120
Db	21	IleAlaAlaGluAspValMetIleValSerAlaSerGlyTyrGluLysLeuThr	40

121 AACGAGCCGCGCTGTTTCTGTTAGTCCAGGAGAAATTCAGTCCAGCCAGTACCAC 180
 41 AsnAlaAlaSerValSerValIleSerGlnGluLeuGlnSerSerGlnTyrHis 60
 181 GATCTGGGGAGGCTCTGAGATCAGTAGAGGGTGTGATTTGAAAGTGTGACGGGTAA 240
 61 AspLeuAlaGluAlaLeuArgSerValGluGlyValAspValGluSerGlyThrGlyLys 80
 241 ACCGAGGCTGGAATCAGCATCCGAGGAGATCCAGCCAGTTACAGCTGATACATGAT 300
 81 ThrGlyGlyLeuGluIleSerIleArgGlyMetProIleSerThrLeuIleLeuIle 100
 301 GATGGTGTCTGAGCGGAGGAGCAGTACGTACCTCCCAACGGTTTTCTGCCATGAAT 360
 101 AspGlyValArgGlnGlySerSerAspValThrProAsnGlyPheSerAlaMetAsn 120
 361 ACCGGTTCATGCCCTCTGCGCCGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTG 420
 121 ThrGlyPheMetProProLeuAlaAlaIleGluArgIleGluValIleArgGlyProMet 140
 421 TCCACACTGTATGCTCTGATGCGATGGCGGCTGTGTGAATATCATACCGAGAAAGAT 480
 141 SerThrLeuTyrGlySerAspAlaMetGlyValValAsnIleIleThrArgLysAsn 160
 481 GCAGACAAATGGCTCTCTTCGTCATGCGAGGCTGAACTGCGAGGAAAGCAAAATGG 540
 161 AlaAspLysTrpLeuSerSerValAsnAlaGlyLeuAsnLeuGlnSerAsnLysTrp 180
 541 GGTAAACAGCAGCAGTAAATTTCTGGAGCAGTGGTCCCTGTGATGATCTGTGACG 600
 181 GlyAsnSerSerGlnPheAsnPheTrpSerSerGlyProLeuValAspSerValSer 200
 601 CTGAGGTACCGGTAGCACACACAGCTCAGGGTTTCATCGGTACATCATCTGACGAT 660
 201 LeuGlnValArgGlySerThrGlnArgGlnGlySerSerValThrSerLeuSerAsp 220
 661 ACAGCAGGCGAGTATCTCTATCCAGGAGTACAGAGTACAGAAATTAATCTTGGTGCAG 720
 221 ThrAlaAlaThrArgIleProTyrProThrGluSerGlnAsnTyrAsnLeuGlyAlaArg 240
 721 CTTCAGTGAAGCGTGGAGCAGGATGTCTCTGTTTATATGATGATACCAACCGGCGAG 780
 241 LeuAspTrpLysAlaSerGluGlnAspValLeuTrpPheAspMetAspThrThrArgGln 260
 781 CGTTATGATACACGGGATGGGCACTGGGAGTCTGACGGGGGATATACCGGACCTG 840
 261 ArgTyrAspAsnArgAspGlyGlnLeuGlySerLeuThrGlyGlyTyrAspArgThrLeu 280
 841 CGCTATGAGCGAAACAAATTTTCAGCTGGCTATGATCATCTTTCACCTTCGGAACATGG 900
 281 ArgTyrGluArgAsnLysIleSerAlaGlyTyrAspHisThrPheThrPheGlyThrTrp 300
 901 AATCGTATCTGAATCGGACGACGACAGAAATAAAGTCTGTCAGCTTGTACGCGAGTGA 960
 301 LysSerTyrLeuAsnTrpAsnGluThrGluAsnLysGlyArgGluLeuValArgSerVal 320
 961 CTGAAGCGGACAAATGGGGCTTGGCTGAGCGCGGAGCTTAAGGAATCGAACCTT 1020
 321 LeuLysArgAspLysTrpGlyLeuAlaGlyGlnProArgGluLeuLysGluSerAsnLeu 340
 1021 ATCTGAATTCATCTGCTTACCTCTGGGAGAAATCTCATCTGTTACGTTGGGGGCG 1080
 341 IleLeuAsnSerLeuLeuLeuThrProLeuGlyGluSerHisLeuValThrValGlyGly 360
 1081 GAGTTTCAGAGCTCGTCCATGAAGACGAGTGTCTTCCAGCAGCAGCGTGAATCTTC 1140
 361 GluPheGlnSerSerSerMetLysAspGlyValValLeuAlaSerThrGlyGluThrPhe 380
 1141 CGGCAGAAAGCTGGTGGTATTTGCTGAGGATGAGTGGCATCTCACGAGTGCATCTCGC 1200
 381 ArgGlnLysSerTrpSerValPheAlaGluAspGluTrpHisLeuThrAspAlaLeuAla 400

1201 CTGACTGCGGCGAGCCGCTATGAACATCATGAGCAATTCGGGGGACACTTCAGTCCGCGT 1260
 401 LeuThrAlaGlySerArgTyrGluHisGluGlnPheGlyGlyHisPheSerProArg 420
 1261 GCATATCTGCTCTGGAGTGTGGCAGATGCTCGGACGCTGAAAGCGGTGTGTACACCGGA 1320
 421 AlaTyrLeuValTrpAspValAlaAspAlaTrpThrLeuLysGlyValThrThrGly 440
 1321 TATAAGGACACACAGATTCGGGCGACTACATAAAGGATTAAGTGTGTCTCCGGCAGGGA 1380
 441 TyrLysAlaProArgMetGlyGlnLeuHisLysGlyIleSerGlyValSerGlyGlnGly 460
 1381 AAAACAAATCTACTTGTGTACCCGACCTGAAGCGGAGAGAGCGTCAGTTATGAGCTT 1440
 461 LysThrAsnLeuLeuGlyAsnProAspLysProGluGluSerValSerTyrGluAla 480
 1441 GGGGTGTATTACGATAACCCCGCGCTCTGAATCCCAATGTCACAGGTTTATGACTGAC 1500
 481 GlyValTyrTyrAspAsnProAlaGlyLeuAsnAlaAsnValThrGlyPheMetThrAsp 500
 1501 TTCTCCAAACAGATTTGCTCTTATTCATATAATGATAACCAATAGCTATGTAAACAGC 1560
 501 PheSerAsnLysIleValSerTyrSerIleAsnAspAsnThrAsnSerTyrValAsnSer 520
 1561 GGAAGGCCCGGTTGCGAGGTGTGGAATTTGCCGGCACATTCGCGTGTGTGTCAGGAT 1620
 521 GlyLysAlaArgLeuHisGlyValGluPheAlaGlyThrLeuProLeuTrpSerGluAsp 540
 1621 GTCACGCTGTCACTGAATTAACCTGACCCGAGTGAACAACTGATGCTGTGATAACAAA 1680
 541 ValThrLeuSerLeuAsnTyrThrTrpArgSerGluGlnArgAspGlyAspAsnLys 560
 1681 GGTGGCGCGCTAGTTATACCCCTGAACACATGTTGAATCGGAACTGAACCTGCGAGATC 1740
 561 GlyAlaProLeuSerTyrThrProGluHisMetValAsnAlaLysLeuAsnTrpGlnIle 580
 1741 ACCGAAGAGGTGGCATCATGCTGGTGGTCCGCTTATCGCGGAAACACACACCTTTTCA 1800
 581 ThrGluGluValAlaSerTrpLeuGlyAlaArgTyrArgGlyLysThrProArgPheThr 600
 1801 CAGAATTATTGTCACCTGAGCGCTGTACAGAAAGATGTATGATGAGAAAGAGAGATAC 1860
 601 GlnAsnTyrSerSerLeuSerAlaValGlnLysLysValTyrAspGluLysGlyGluTyr 620
 1861 CTGAAGCCTCGGAGGTGGTGGATGCGAGTCTCTGTTGAAGATGACGATGCGCTGACG 1920
 621 LeuLysAlaTrpThrValValAspAlaGlyLeuSerTrpLysMetThrAspAlaLeuThr 640
 1921 CTGAATCTCGGTGAATAACCTGCTCAACAGGATTACAGTACGCTGAGCCTGTACAGT 1980
 641 LeuAsnAlaAlaValAsnAsnLeuLeuAsnLysAspTyrSerAspValSerLeuTyrSer 660
 1981 GCGGTGAAGTAGCTGTATGCGGTGATTCCTTCAGACGGGATCATCAACACAGGA 2040
 661 AlaGlyLysSerThrLeuTyrAlaGlyAspTyrPheGlnThrGlySerThrThrGly 680
 2041 TATGTGATCTCAGCGCAAAATTAATCTGATGCTCGCTGAACCTATCATCAGTTC 2088
 681 TyrValIleProGluArgAsnTyrTrpMetSerLeuAsnTyrGlnPhe 696

RESULT 2
 G85607
 probable receptor Z1178 [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
 C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence_revision 16-feb-2001 #text_change 14-Sep-2001
 C:Accession: G85607; H85658
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,
 Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: A85480; PMID:21074935; PMID:11206551
 A:Accession: G85607
 A>Status: preliminary

A:Molecule type: DNA	
A:Residues: 1-696 <STO>	
A:CROSS-references: GB:AB005174; NID:gl2513986; PIDN:AAG55323.1; GSPDB:GN00145; UWGP:Z11	
A:Experimental source: strain O157:H7, substrain EDL933	
A:Accession: H85658	
A:Status: preliminary	
A:Molecule type: DNA	
A:Residues: 1-696 <ST2>	
A:CROSS-references: GB:AB005174; NID:gl2514503; PIDN:AAG55732.1; GSPDB:GN00145; UWGP:Z16	
A:Experimental source: strain O157:H7, substrain EDL933	
C:Genetics:	
A:Gene: Z1178; Z1617	
C:Superfamily: ferrienterochelin receptor; tonB-dependent receptor amino-terminal homolog	
Alignment Scores:	
Pred. No.:	7,896-242
Score:	2632.00
Percent Similarity:	99.86%
Best Local Similarity:	99.86%
Query Match:	97.32%
DB:	2
US-10-625-972-4 (1-2091) x G85607 (1-696)	
QY	1 ATGCGAATACACACTCTGGCTTCCTAGTATCCCTGCTCGGATTTTCAGCCAGCAGC 60
DB	1 MetArglleThrLeuAlaSerValValilleProCysLeuGlyPheSerAlaSerSer 20
QY	61 ATAGTCTGTCAGAGAGTGTGATGTTGCTCGGCATCCGGCTATGAGAAAGCTGACT 120
DB	21 lleAlaAlaAlaGluAspValMetlleValSerAlaSerGlyTyrgluylsleuThr 40
QY	121 AACGAGCGCGCAGTGTCTGTGATAGCCAGAGGAGATTCAGTCAGTCAGCCAGTACCAC 180
DB	41 AsnAlaAlaAlaSerValilleSerGlnGluGluLeuGlnSerSerGlnTyrrHis 60
QY	181 GATCTGGCGGAGCTCTGAGATCAGTACGAGGTGTGGATGTTGAAGTGTACGGGTAAA 240
DB	61 AspLeuAlaGluAlaLeuArgSerValGluGlyValAspValGluSerGlyThrGlylys 80
QY	241 ACCGAGGGCTGGAATCAGCATCCGAGGAATCCAGCCAGTTACACCTGATCTGATT 300
DB	81 ThrGlyGlyLeuGluileSerilleArgGlyMetProAlaSerTyrrThrLeuileLeuile 100
QY	301 GATGTGTCTCAGGGCGAGCAGTACGTCGACTCCCAAGGTTTTCGCCATGAT 360
DB	101 AspGlyValArgGlnGlySerSerAspValThrProAsnGlyPheSerAlaMetAsn 120
QY	361 ACCGGTTTCATGCCCTCTGCCGCAATTGAGCGTATTGAGGTTATCAGGGGGCCGATG 420
DB	121 ThrGlyPheMetProProleuAlaAlalleGluArglleGluValilleArgGlyProMet 140
QY	421 TCACACTGTATGGCTCTGATCGCATGGCGGTGGTGGTGAATATCATACGAAAGAAAT 480
DB	141 SerThrLeuTyrrGlySerAspAlaMetGlyGlyValValAlaSerilleThrArglysAsn 160
QY	481 GCAGCAAAATGGCTCTCTTCCTCAATGACAGGCTGAATCTCAGGAAGCAACAAATGG 540
DB	161 AlaAspLysTrpLeuSerSerValAsnAlaGlyLeuAsnLeuGlnGluSerAsnLysTrp 180
QY	541 GGTAAACAGCAGCCAGTTTAAATTTCTGGACAGTGGTCCCTTCTGGATGATTTCTGACG 600
DB	181 GlyAsnSerSerGlnPheAsnPheTrpSerSerGlyProLeuValAspAspSerValSer 200
QY	601 CTCAGGTACGGGTAGCACACACAGCGTCAGGTTTCATCGGTACATCAGTCAGCGAT 660
DB	201 LeuGlnValArgGlySerThrGlnGlnArgGlnGlySerSerValThrSerLeuSerAsp 220
QY	661 ACAGCAGGACCGGTATCTCTATCCACGAGTACAGAAATTAATCTTGTGTGACGT 720
DB	221 ThrAlaAlaThrArgilleProTyrrProThrGluSerGlnAsnTyrrAsnLeuGlyAlaArg 240
QY	721 CTTGATCGGAAGCGCTCGGAGCAGGATGTCTCTGGTTTGTATGATACCCCGGCAG 780

DB	241 LeuAspTrpLysAlaSerGluGlnAspValLeuTrpPheAspMetAspThrThrArgGln 260
QY	781 CGTTATGATAACCGGATCGGCAACTGGGAGTCTCACGGGGGATATGACCGGACCCTG 840
DB	261 ArgTyrrAspAsnArgAspGlyGlnLeuGlySerLeuThrGlyGlyTyrrAspArgThrLeu 280
QY	841 CGTATGAGCGAAACAAAATTTTCAGTGGCTGATGATCATCTTCACTTCCGGAACATGG 900
DB	281 ArgTyrrGluArgAsnLysilleSerAlaGlyTyrrAspHisThrPheThrPheGlyThrTrp 300
QY	901 AAATCGTATCTGAACCTGGACGACGACAGAAAATAAAGGTCTGAGCTTGTACCGAGTGA 960
DB	301 LysSerTyrrLeuAsnTrpAsnGlnThrGluAsnLysGlyArgGluLeuValArgSerVal 320
QY	961 CTGAAGCGGACAAATGGGGCTTGGCGTACGCCGGAGCTTAAAGAACTGCAACCTT 1020
DB	321 LeuLysArgAspLysTrpGlyLeuAlaGlyGlnProArgGluLeuLysGluSerAsnLeu 340
QY	1021 ATCTGATTTCACTACTGCTTACCTCTCGGAGAACTCATCTGTTTACGGTGGGGGC 1080
DB	341 lleLeuAsnSerLeuLeuLeuThrProLeuGlyGluSerHisLeuValThrValGlyGly 360
QY	1081 GAGTTTCAGAGCTCGTCCATGAAAGACGAGGATTTGCTTTGCCACGACAGGTGAACCTTC 1140
DB	361 GluPheGlnSerSerSerMetLysAspGlyValValLeuAlaSerThrGlyGluThrPhe 380
QY	1141 CGGCAAAAAGCTGGTTCGTTATTTCTGAGGATGAGTGGCATCTCAGGATGACATTGG 1200
DB	381 ArgGlnLysSerTrpSerValPheAlaGluAspGluTrpHisLeuThrAspAlaLeuAla 400
QY	1201 CTGACTCGGCGACGCGCTATGAACATCATGACCAATTCGGGGGACACTTCAGTCCGCGT 1260
DB	401 LeuThrAlaGlySerArgTyrrGluHisGluGlnPheGlyGlyHisPheSerProArg 420
QY	1261 GCATATCTGCTCGGATGTGGCAGATGCTCGAGCTGAAAGCGGTGTGACACGGGA 1320
DB	421 AlaTyrrLeuValTrpAspValAlaAspAlaTrpThrLeuLysGlyGlyValThrThrGly 440
QY	1321 TATTAAGCACCCAGATGGGCGAGTACATAAAGGATTAAGTGTGTGTCGGGCGAGGA 1380
DB	441 TyrrLysAlaProArgMetGlyGlnLeuHisGlyGlylleSerGlyValSerGlyGlnGly 460
QY	1381 AAAACAAATCTACTTGTGTAACCCCGACCTGAAGCCGGAAGAGCGCTCATGAGGCT 1440
DB	461 LysThrAsnLeuLeuGlyAsnProAspLeuLysProGluGluSerValSerTyrrGluAla 480
QY	1441 GGGGTGATTAACGATAACCCCGCGCTCAATGCCAATCTCACAGGTTTATGACTGAC 1500
DB	481 GlyValTyrrAspAsnProAlaGlyLeuAsnAlaAsnValThrGlyPheMetThrAsp 500
QY	1501 TTCTCCCAACAGATTGCTCTTATTCATAAAATGATAACCACTAGCTATGTAAACAGC 1560
DB	501 PheSerAsnLysilleValSerTyrrSerilleAsnAspAsnThrAsnSerTyrrValAsnSer 520
QY	1561 GGAAGAGCCCGGTTGACCGGTGGAAATTTCCGGGCACATTCGCGCTGTGGTCAAGAT 1620
DB	521 GlyLysAlaArgLeuHisGlyValGluPheAlaGlyThrLeuProLeuTrpSerGluAsp 540
QY	1621 GTCAAGCTGTACATGATTAACCTGACCCGAGTCAACCACTGATGGTGTAAACAAA 1680
DB	541 ValThrLeuSerLeuAsnTyrrTrpThrArgSerGluGlnArgAspGlyAspAsnLys 560
QY	1681 GGTGGCGCTGAGTTATACCCCTGAACACATGTTGAATGCAAACTGAACTGCAGATC 1740
DB	561 GlyAlaProLeuSerTyrrThrProGluHisMetValAsnAlaLysLeuAsnTrpGlnIle 580
QY	1741 ACCGAAGGTGGCATCATGGCTGGTCCCGTTATCGCGGAAACACACAGCTTTCACC 1800
DB	581 ThrGluGluValAlaSerTrpLeuGlyAlaArgTyrrArgGlyLysTrpProArgPheThr 600
QY	1801 CAGAAATTTCTGCTCACTGAGCGCTGTACAGAAAGAGTGTATGATGAGAAAGGAGATAC 1860


```

:::
467 ArgPheProThrTyrGlyAsnProAspLeuLysGluThrSerLeuAsnTyrGluIle 486
1441 GGGGTGTTATAGCAAAACCCCGCGTCTGAATGCCAATGTACAGGTTTATGACTGAC 1500
487 AlaAlaIleTyrAsnAsnAspLeu---PheTyrValSerAlaThrGlyPheLeuThrAsn 505
1501 TTCTCCAAACAAGATTGCTCTTATTCATAAT----- 1533
506 PheLysAspLysIleSerSerGlnSerTyrAsnSerGluProIleProGlyIleGly 525
1534 -----GATAACACCAATAGCTATTAACACGGGAAGCCCGGTGACCGGT 1581
526 ThrCysAspAlaAspArgCysSerArgAlaIleAsnHisGlyLysValGluTyrLysGly 545
1582 GTGGGAATTT---GCCGGCACAATGCGCTGTGTGTCAGAGAGTGCACGCTCTCTACTGAAT 1638
546 ValGluLeuGlyAlaGlyIleSerProLeu-----AspAsnLeuAsnValAsnPheAla 563
1639 TACACTGGACCCGAAGTGAACAAGTGAATGGTGTATGATACAA-----GGTCCGCG 1689
564 TyrThrTyrLeuAspThrGluValLysGluAlaGlnAspArgSerValIleGlyLysPro 583
1690 CTGAGTTATACCCCTCAACACATGTGAATGCGAACTGAACCTGGCAGATCACCGAAGAG 1749
584 GluGlnAspSerLeuLysHisAsnIleMetLeuLysThrGluTyrSerPheTyrAsnLys 603
1750 GTGGCATCATGCTCGGTGCC-----CGTTATCGCGGGAACACCCAGT 1794
604 IleThrProTyrIleLysGlyGluTyrGlnIleAspArgTyrMetGlyAspThr----- 621
1795 TTCACCCAGAATATTCGTCACCTGAGCGGTGTACAGAGAAAGTGTATGATGAGAAAGGA 1854
622 -----AsnIleAsnArg 625
1855 GAATACCTGAAGCCCTGGACGGTGTGGATGACGAGTCTGCTGGAAAGATACCGATGCC 1914
626 GluTyrTyrLysAspIlePheLeuAlaSerMetGlyValArgTyrAspIleAsnLysGln 645
1915 CTGACGCTGAATGCTCGGTGAATACCTGCTCAACAGGATTAACAGTAC---GTGAGC 1971
646 TrpSerIleAsnAlaAlaIleTyrAsnLeuPheAspLysSerPheThrAsnGlyTrpGlu 665
1972 CTGTACAGTCCCGTAAAGTAGTACGCTGTATGCGGTGATTAATCTCCAGACGGGATCATCA 2031
666 SerTyrAlaSerGlySerGlySerThrTrpValAsnThrTyr----- 679
2032 ACAACAGGATATGTGATCTACGCGAATTAATCTGATCTCGCTGAACTACATCAGTTC 2088
680 -----AsnArgIleGluGluGlyArgArgMetTyrIleSerIleAsnGlyAsnPhe 696

RESULT 4
D82317
iron-regulated outer membrane virulence protein, TonB receptor family VC0475 [imported]
C;Species: Vibrio cholerae
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C;Accession: D82317
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
Chardson, D.; Mekalanos, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, B.
I, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A;Reference number: AB2035; MUID:20406833; PMID:10952301
A;Accession: D82317
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-652 <HEI>
A;Cross-references: GB:AB041134; GB:AE003852; NID:9954900; PIDN:AAF93648.1; GSPDB:GN001
A;Experimental source: serogroup O1; strain N16961; biotype El Tor
C;Genetics:
A;Gene: VC0475
A;Map position: 1
C;Superfamily: ferrienterochelin receptor; tonB-dependent receptor amino-terminal homolo

```

```

Alignment Scores: 1.74e-64 Length: 652
Pred. No.: 1052.50 Matches: 258
Score: 51.80% Conservative: 116
Percent Similarity: 35.73% Mismatches: 223
Best Local Similarity: 28.20% Indels: 125
Query Match: 2 Gaps: 24
DB:

US-10-625-972-4 (1-2091) x D82317 (1-652)
QY 40 CTGGGATTTTACGCCAGCAGCATAGCT-----GCTCCAGAGATGTGATGATT 87
DB 17 LeuMetPheSerAlaSerAlaPheAlaGlnAspAlaThrLysThrAspGluThrMetVal 36
QY 88 GTCTCGCATCCGCTATGAGAAAAGCTACTAAGCCAGCCGCCAGTCTTCTGTGATT 147
DB 37 ValThrAlaAlaGlyTyrAlaGlnValIleGlnAsnAlaProAlaSerIleSerValIle 56
QY 148 AGCCAGAGAAATGGCAGTCCAGCCAGTACACGATCTGCGGAGGCTCTGAGATCAGTA 207
DB 57 SerArgGluAspLeuGluSerArgTyrTyrArgAspValThrAspAlaLeuLysSerVal 76
QY 208 GAGGGTGTGATGTTGAAAGTGTACGGGTAAACCCGAGGGCTGGAATTCAGATCCGA 267
DB 77 ProGlyValThrValThrGlyGlyGlyAspThrThr-----AspIleSerIleArg 93
QY 268 GGAATGCCAGCCAGTTACAGCTGATATGATTGATGTTGTTCTGTCAGGCGGAGAGCAGT 327
DB 94 GlyMetGlySerAsnTyrThrLeuIleValAspGlyLysArgGln---ThrSerArg 112
QY 328 GACGTGACTCCCAACGGTTTTCT---GCCATGAATACCGGGTTCATGCCCTCTCGCC 384
DB 113 GlnThrArgProAsnSerAspGlyProGlyIleGluGlnGlyTrpLeuProLeuGln 132
QY 385 GCCATTGAGGCTATTAGGTATTCAGGCGCCGATGTCACATGATGCTCTGTGATGCG 444
DB 133 AlaIleGluArgIleGluValIleArgGlyProMetSerThrLeuTyrGlySerAspAla 152
QY 445 ATGGCGGTGTGGTGAATATCATTACAGAAAGAACATGCAGACAAATCGCTCTCTCGTC 504
DB 153 IleGlyGlyValIleAsnIleIleThrArgLysAspGlnGlnTrpSerGlyAsnVal 172
QY 505 AATCCAGCGGTGAATCTCGAGGAAAGCAACAATGGGGTAACACGACCCAGTTTAATTC 564
DB 173 GlnLeuSerThrValValGlnGluAsnArgAlaSerGlyAspGluGlnSerAlaAsnPhe 192
QY 565 TGGAGCAGTGTCCCTTGTGATGATTCTGCTGAGCTGACGCTGACGGTAGCACACAA 624
DB 193 PheValThrGlyProLeu---SerAspAlaLeuSerLeuGlnValTyrGlyGlnThr 211
QY 625 CAGCGTTCAG-----GGTTCATCGTCACTCATCATCTGAGC 657
DB 212 GlnArgAspGluAspGluIleGluHisGlyTyrGlyAspLysSerLeuArgSerLeu--- 230
QY 658 GATACAGCAGCCAGCGGTATCTCTTATCCACGAGTCAACAGATTAATCTTGGT--- 714
DB 231 -----ThrSerLysLeuAsnTyrGlnLeuAsnPro 240
QY 715 -----GCACGTCTTACTGGAAGCGCTGCGAGCAGGATGTGCTCTGTTTGAATG 765
DB 241 AspHisGlnLeuGlnLeuGluAlaGlyValSerAlaGlnAsp----- 254
QY 766 GATACACCCCGCAGCGTTATGATAACCGGATGGGCACTGGGAGTCTCAGCGGGGGA 825
DB 255 -----ArgGluAsnAsnValGlyLysSerAlaGlnSerSerGlyCysArgGlyThr 271
QY 826 TATGACCGGACC---CTGCGCTATGAGCGAAACAAATTTTCAGCTGGCTATGATCATACT 882
DB 272 CysSerAsnThrAspAsnGlnTyrArgArgAsnHisValAlaValSerHisGln----- 289
QY 883 TTCACCTTCGGAACATGGAAA-----TCGTATCTGAATCGAACGAG 924
DB :::::

```


290 -----GlyAspTrpGlnAspValGlyGlnSerAspThrThrLeuGlnThrGluGlu 306
 925 ACAGAAATAAGGCTGAGCTGTACGAGTGTACTAGAGCGGACAAATGGGGCTT 984
 307 AsnThrAsnLysSer----- 311
 985 GCCGGTCAGCGCGGAGCTTAAGAAATCGAATCTATCTGAAATTCATTACTGCTTACC 1044
 312 -----AsgGluMetSerIleAspAsnThrValPheLysSerThrLeuValAla 327
 1045 CTTCTGGGAGATCTCATCTGTTACGCTGGGGGGGAGTTTCAGAGCTGCTCCATGAA 1104
 328 ProIleGlyGlu---HisMetLeuSerPheGlyValGlyLysHisGluSerLeuGlu 346
 1105 GACGGAGTTGCTTCGCCAGCAGAGTGAACACT---TTCGGCGAGAAAGCTGGTGGTA 1161
 347 AspLysThrSerAsnLysIleSerSerArgThrHisIleSerAsnThrGlnTrpAlaGly 366
 1162 TTTCTGAGGATGAGTGGCATCTCAGGATGCACTTGGCTGCTGCTGCGGAGCGCTAT 1221
 367 PheIleGluAspGlnTrpAlaLeuAlaGluGlnPheArgLeuThrPheGlyGlyArgLeu 386
 1222 GAACATCATCAAGAAATTCGGGGGACACTTCAGTCCGCGTGCATATCTGCTGGGATG 1281
 387 AspHisAspLysAsnThrGlySerHisPheSerProArgValThrGlyValTrpAsnLeu 406
 1282 GCAGATCGCTGAGCGTGAAGCGGTGTGACACCGGATATAAGACCCAGAAATGGGG 1341
 407 AspProLeuThrThrValLysGlyValSerThrGlyPheArgAlaProGlnLeuArg 426
 1342 CAGCTACATAAGGATTAGTGTGTGTCGGGCGAGGAAACAAATCTACTTGGTAC 1401
 427 GluValThrProAspTrpGlyGlnValSerGlyGlyGly-----AsnIleThrGlyAsn 444
 1402 CCCGACTGAGCGGAGAGCGCTGATGAGCTGGGTGTATACGATACCC 1461
 445 ProAspLeuLysProGlnThrSerIleAsnLysGluLeuSerLeuMetThrThrGly 464
 1462 GCCGCTGTAATGCAATGTCAGAGTTTATGACTGACTTCCAAAGATTTGCTCT 1521
 465 SerGlyLeuAlaAlaSerLeuThrAlaPheHisAsnAspPheLysAspLysIleThrArg 484
 1522 TATTCATAAATGATAAC-----ACCAATAGCTAT----- 1551
 485 ValAlaCysProAlaAsnIleCysThrAlaGlyProAsnGlnTrpGlyAlaAlaProThr 504
 1552 -----GTAAACGCGAAGCGCGGTGCGAGTGTGGAATTTCCGCGCATTCGCG 1605
 505 TyrArgValAsnIleAspGluAlaGluThrThrGlyAlaGluAlaThrLeuSerLeuPro 524
 1606 CTGTGTCAGAGCATGTCAGCTGTGCTACATGATTTACACCTGGACCGAGTCAACAAGT 1665
 525 Ile---ThrGluSerValGluLeuSerSerThrThrHisSerGluGlnLys 543
 1666 GATGGTGAATAAAGGTCGCGCTGAGTTATACCCCTGAACACATGGTGAATCGGAA 1725
 544 SerGlyAsnPheAlaGlyArgProLeuLeuGlnLeuProLysHisLeuPheAsnAlaAsn 563
 1726 CTGAATGCGAGATCAACGAGAGTGGCATCATGCTGGTGGTGGTGGTGGTGGTGGTGG 1785
 564 LeuSerTrpGlnThrThrAspArgLeuAsnSerTrpAlaAsnLeuAsnThrArgGlyLys 583
 1786 ACACCAGCTTTCCACCAGAAATTTATTCCTGCTGAGCGCTGTACAGAAAGATGTATGAT 1845
 584 -----GluMetGlnProGluGlyGlyAla 591
 1846 GAGAAAGGAGATACCTG---AAAGCTGAGCGGTGGTGGATGAGCTGCTGCTGGAG 1902
 592 SerAsnAspPheIleAlaProSerThrPheIleAspThrGlyValThrTrpAla 611
 1903 ATGACGGATGCGCTGAGTGAATGCTGGGTGAATAACCTCTCAACAGGATTTACAGT 1962
 612 LeuThrAspThrAlaThrIleLysAlaAlaValTyrAsnLeuPheAspGlnGluValAsn 631

QY 1963 GACGTGAGCTGTACAGTCCGGTAAGAGTAGTACGCTGTATCCGGTGTACTTCCAC
 Db 632 -----TyrAla---GluTyr----- 635
 QY 2023 GGATCATCAACAACAGGATATGTATACCTGAGCAAAATTTACTGATGTCGCTGAATAT 2082
 Db 636 -----GlyTyrValGluAspGlyArgGlyTyrTrpLeuGlyLeuAspIle 650
 QY 2083 CAGTTC 2088
 Db 651 AlaPhe 652
 RESULT 5
 S25265
 outer membrane protein irga precursor - Vibrio cholerae
 N:Alternate names: ferrienterochelin receptor homolog
 C:Species: Vibrio cholerae
 C:Date: 28-May-1993 #sequence_revision 28-May-1993 #text_change 08-Oct-1999
 C:Accession: S25265; A37834
 R:Goldberg, M.B.; Boyko, S.A.; Butterton, J.R.; Stoeber, J.A.; Payne, S.M.; Calderwood, M.O. Microbiol. 6, 2407-2418, 1992
 A:Title: Characterization of a Vibrio cholerae virulence factor homologous to the family
 A:Reference number: S25265; MUID:93023868; PMID:1406279
 A:Accession: S25265
 A:Molecule type: DNA
 A:Residues: 1-652 <GOL>
 A:Cross-references: GB:U72152; EMBL:M63192; NID:G1763224; PIDN:AAC44766.1; PID:G1763226
 A:Note: the sequence from Fig. 3 is inconsistent with that from Fig. 2 in having 299-Thr
 R:Goldberg, M.B.; Boyko, S.A.; Calderwood, S.B.
 J. Bacteriol. 172, 6863-6870, 1990
 A:Title: Transcriptional regulation by iron of a Vibrio cholerae virulence gene and homo
 A:Reference number: A37834; MUID:91072235; PMID:2174861
 A:Accession: A37834
 A:Molecule type: DNA
 A:Residues: 1-152, 'D', <GO2>
 A:Cross-references: GB:M37773
 C:Genetics:
 C:Superfamily: ferrienterochelin receptor; tonB-dependent receptor amino-terminal homolog
 C:Keywords: membrane protein
 F:1-25/Domain: signal sequence #status predicted <SIG>
 F:26-652/Product: outer membrane protein irga #status predicted <MAT>
 F:68-214/Domain: tonB-dependent receptor amino-terminal homology <TNN>
 F:367-652/Domain: tonB-dependent receptor carboxyl-terminal homology <TNC>
 Alignment Scores:
 Pred. No.: 1,74e-64 Length: 652
 Score: 1052.50 Matches: 258
 Percent Similarity: 51.94% Conservative: 117
 Best Local Similarity: 35.73% Mismatches: 222
 Query Match: 28.20% Indels: 125
 Db: 24 Gaps:
 US-10-625-972-4 (1-2091) x S25265 (1-652)
 QY 40 CTCGGATTTTCAGCCAGCAGCATAGCT-----GCTCGAGGAGTGTGATGATT 87
 Db 17 LeuMetPheSerAlaSerAlaPheAlaGlnAspAlaThrLysThrAspGluThrMetVal 36
 QY 88 GTCTCGCATCCGGCTATGAGAAAAAGTACTGACTACCGCCAGCCGCTGTTCTGTGATT 147
 Db 37 ValThrAlaAlaGlyTyrAlaGlnValIleGlnAsnAlaProAlaSerIleSerValIle 56
 QY 148 AGCCAGGAGAAATTCAGTCCAGCCAGTACCCAGTCCGCGAGGCTCTGAGATCAGTA 207
 Db 57 SerArgGluAspLeuGluSerArgTyrTyrArgAspValThrAspAlaLeuLysSerVal 76
 QY 208 GAGGGTGTGATGTTGAAAGTGGTACGGGTAAACCCGAGGGCTGGAAATCAGCATCCGA 267
 Db 77 ProGlyValThrValThrGlyGlyGlyAspThrThr-----AspIleSerIleArg 93
 QY 268 GGAATGCCAGCCAGTTACACGCTGATGATTGATTGTTGCTCGTCAGGCGGAGCAGT 327

Db 607 -----MetGlySerArgThrProSerTyrThrPheValAspLeu 619
 Qy 1888 GGTCTGTCGTGAAGATACGCGATGCCCTGACGTGAATGCTGGGTGATATACCTGCTC 1947
 Db 620 GlyAlaAsnTyrGlnLeuThrLysGluValArgLeuMetGlyGlyValTyrAsnLeuLeu 639
 Qy 1948 AACAG-----GATTACAGTCAGCTGAGCCTGTACAGTCCGGGTAAAGTAGC 1995
 Db 640 AspLysArgValAspIleAspValAsnAsp----- 649
 Qy 1996 CTGTATGCCGGTGATTAATCCAGACGGGATCATCAACACAGGATATGTCATACCTGAG 2055
 Db 650 -----LysValLeuAspGly 654
 Qy 2056 CGAATTAATCTGATGCTGCTGAACATACAGTTC 2088
 Db 655 ArgArgTyrMetValGlyAlaSerTyrAspPhe 665
 RESULT 7
 AF3038
 exogenous ferric siderophore receptor bfrA [imported] - Agrobacterium tumefaciens (strain
 C)Species: Agrobacterium tumefaciens
 C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
 C/Accession: AF3038
 R/Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell
 ; Karp, P.; Romero, P.; Zhang, S.
 Science 294, 2317-2323, 2001
 A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
 ster, E.W.
 A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
 A/Reference number: AB2577; MUID:21608550; PMID:11743193
 A/Accession: AF3038
 A/Status: Preliminary
 A/Molecule type: DNA
 A/Restrictions: 1-702 <KUR>
 A/Cross-references: GB:AE008689; PID:AL44724.1; PID:gl7742356; GSPDB:GN00187
 A/Experimental source: strain C58 (Dupont)
 C/Genetics:
 A/Gene: bfrA
 A/Map position: linear chromosome
 A/Superfamily: ferrienterochelin receptor; tonB-dependent receptor amino-terminal homolog

Alignment Scores:
 Pred. No.: 2,17e-60 Length: 702
 Score: 993.00 Matches: 242
 Percent Similarity: 49.80% Conservative: 133
 Best Local Similarity: 32.14% Mismatches: 244
 Query Match: 26.61% Indels: 134
 DB: 2 Gaps: 23

US-10-625-972-4 (1-2091) x AF3038 (1-702)
 Qy 19 GCTTCGTAGTATCCCTGCTCGGATTTTCAGCCAGCAGCATAGCT-----GCTGCA 72
 Db 21 AlaAlaMetAlaGlyThrAlaLeuGlyPheAlaValProAlaPheAlaGlnAlaSer 40
 Qy 73 GAG-----GATGTGATGTCTCGCATCCGGATCCGGTATGAGAAAAGCTG 117
 Db 41 GluGlyAsnThrValLeuGlnGlnIleValThrAlaSerGlyPheGlnGlnAsnVal 60
 Qy 118 ACTAACGCCGCCAGTCTTCTGTGATTAGCCAGGAGGATTCAGTCCAGCCAGTAC 177
 Db 61 LysAspAlaProAlaSerIleThrValValThrArgGluAspLeuGlnLysGlySerTyr 80
 Qy 178 CAGCATCTCGCGGAGCTCTGAGATCAGTAGAGGTGTGGATGTGTAAGTGTGTCGGGT 237
 Db 81 ArgAspLeuThrAspAlaLeuArgGluValGlnGlyValSerVal-----ThrGly 97
 Qy 238 AAAACCGGAGGGCTGGAATACGATCCGAGGAATGCCAGCAGTATACCGTGTACTG 297
 Db 98 IleAlaAsnGluLysAspValPheIleArgGlyLeuProGlyAlaTyrThrLeuIleLeu 117

Qy 298 ATTGATGGTGTTCGTGAGGCGGAGAGCAGTGCAGTCTGACTCCCAACGGTTTTTCTGCGATC 357
 Db 118 ValAspGlyValArgGln---SerThrArgAspAlaArgThrAsnGlyAsnSerGlyPhe 136
 Qy 358 AATACCGGTTTCATGCCCCCTCTGGCCGCATTCAGCGTATGAGTTATCAGGGGCGG 417
 Db 137 GluGlnSerPheValProValSerAlaIleGluArgIleGluValValArgGlyPro 156
 Qy 418 ATGTCACACTGTATGGTCTGATGCGATGGGCGGTGTGTGAATATCATACCAAGAG 477
 Db 157 MetSerSerLeuTyrGlySerAspAlaMetGlyGlyValIleAsnIleIleThrArgLys 176
 Qy 478 AATGCAGACAAATGGTCTCTTCCTGCAATGAGGGCTGAATCTGCAGGAACACAA 537
 Db 177 ValGlyAspValTyrSerGlySerValThrThrGluGlyThrValGlnGlnHisSerLys 196
 Qy 538 TGGGGTAACAGCAGCCAGTTCATTTCTGAGCAGTGTCCCTCTGATGATTCCTGTC 597
 Db 197 PheGlyAsnSerGlyGlnValSerTyrTyrAlaAsnGlyProIleLeuLysAspGlnLeu 216
 Qy 598 AGCTTCGAGTACGCGGTAGCACACACAGCGCTCAGGGTTCATCGGTACATCCTGAGC 657
 Db 217 GlyLeuGlnLeuTyrGlyArgGlyPheThrArgGlyGluAspArgIle-----LeuAsn 234
 Qy 658 GATACAGCAGCCAGGTATTCCTTATCCACGAGTCCACAGATTCATATATCTTGTGCA 717
 Db 235 GlyThrThrGly-----AlaLysGluTyrAspPheAsnGly 246
 Qy 718 CGTCTGACATGGAAGGCGTCGAGCAGGATGTCTCTGTTTGTATGATGATACACCCGG 777
 Db 247 ArgLeuThrPheThrProAsnGluAspHisAspIleTyrLeuGluGlyGlyLysThrArg 266
 Qy 778 CAGCGTTATGATAACCGGATGGCAACTGCGGAGTCTGACGGGGGATATGACCGGACC 837
 Db 267 LeuArg-----ArgAspAlaGluProGlyAspThrLeuAlaAlaThrAspAlaAsn 283
 Qy 838 CTGCGCTATGAGCGAAC-----AAAATTTCAGCTGGCTATGATCATCTTTCACTTCGGA 894
 Db 284 GlyThrTyrAsnThrAsnThrArgAspHisTyrSerLeuSerHisThrGlyArgTyrGly 303
 Qy 895 -----ACATGGAATCGTATCTGAATCTGAACTCGAAGCAG 924
 Db 304 ProThrThrSerGluPheSerPheGlnGlnGluTyrAlaGluArgThrAsnPheThrArg 323
 Qy 925 ACAGAAATAAAGTCTGAGCTTGTCAGCAGTGTACTGAGCGCGCAATGGGGCTT 984
 Db 324 AsnThrArgThrGlyArg-----Val 330
 Qy 985 GCCGTCAGCGCGGAGCTTAAGGAATCGAATCTTATCTGAAATTCATTACTGCTTACC 1044
 Db 331 ThrGluAsnProArgSerProGluIleArgAsnThrValLeuAspGlyLysPheThrThr 350
 Qy 1045 CCT-----CTGGAGAAATCTCATCTCGTTACGGTGGGGGGGAGTTTCAGAGCTCGTCC 1098
 Db 351 PropheGluLeuPheGlyAsnHisThrLeuValThrGlyGlyGlnTyrPheGluAlaArg 370
 Qy 1099 ATGAAACACCGGAGTGTCTTCCTCCAGCACA-----GCTCAAACTTTCCGCGCAAAAAGC 1152
 Db 371 LeuThrAspGlnAsnProGlyArgArgThrAspArgAspGluThrPheSerAlaThrGln 390
 Qy 1153 TGCTCGGTATTTGCTGAGGATGATGGCATCTCCAGCATTCACCTGCGCTGACTCGCGGC 1212
 Db 391 TrpAlaLeuPheLeuGluAspGluTyrArgIleValAspAsnPheAlaLeuThrGlyGly 410
 Qy 1213 AGCGCTATGACATCATGAGCAATTCGGGGGACATTCAGTCCGCTGATATCTCGTC 1272
 Db 411 LeuArgLeuAspAsnHisGluLysTyrGlyAsnHisPheSerProArgLeuTyrGlyVal 430
 Qy 1273 TGGCATGTCGAGATGCTGACGCTGAAAGCGGTGTGACCCAGCGGATATAAGCGACCC 1332
 Db 431 TrpSerAlaThrGluGluLeuThrIleLysGlyValSerThrGlyPheArgAlaPro 450
 Qy 1333 AGAATGGGGCAGCTACATATAAGGGATTAGTGTGTCTCCGGGCGGCAAAAACAAAT--- 1389

```

Db 451 GluileargGlnleAlaProGlyTyrAlaTyrThrThrGlyGlyGlyCysSerTyr 470
Qy 1390 -----CTACTTGGTAACCCGACCTGAAGCCGGAAGAGAC 1425
Db 471 GlyProSerGlyThrCysGlyValIleGlyAspProAsnLeuGluAlaGlySer 490
Qy 1426 GTCAATTATGAGCTGGGTGTATTACGATAACCCCGGCTGAATGCCAATGTCA 1485
Db 491 ThrSerTyrGluLeuAlaLeuTyrAspAsn---GlyAspIleAlaLeuGlyAlaThr 509
Qy 1486 GGTTTTATGACTACTTCTCCAAACAAGATT----- 1515
Db 510 TyrPheTyrThrAspPheLysAspLysIleSerAsnAlaLeuValLeuAsnProAspGly 529
Qy 1516 -----GTCTCTTATTCCTATAAT 1533
Db 530 ThrProAlaArgTyrSerGluAspArgAsnTyrArgLeuTyrAsnTyrAsnIleAsp 549
Qy 1534 GATAACACCAATAGTATGTAACACAGCGGAAGCCCGGTGCACGGTGTGAATTTGCC 1593
Db 550 Asp-----AlaIleIleGlnGlyValGluLeuThr 559
Qy 1594 GGCACATTTGCCGTGTGG-----TCAGAGGATGTCCACGCTGTCACTGAATACACCTGG 1647
Db 560 AlaThr-----TyrTyrAlaThrSerGluLeuThrLeuArgGlyAsnTyrThrTyr 576
Qy 1648 ACCGAGTGAACAACGCTGATGATACCAAGGTCCCGCTGAGTATACCCCTGAA 1707
Db 577 ThrHisSerGlnGlnThrGlyAspTyrGluGlyPheProLeuAlaArgThrProGlu 596
Qy 1708 CATATGTGAATCGAACTGAACTGGCAGATCAC-----GAAGAGGTGGCATCATGG 1761
Db 597 HisMetAlaAsnLeuArgGlyAspTyrPheProIleAspGlyLeuGluAlaTyr 615
Qy 1762 CTGGTGGCCGTTATCGCGGAACA-----CCAGTTTACCAGAAATAT 1809
Db 616 ValSerLeuAsnTyrHisGlySerGluIleAsnAlaGlyProArgIleGlyAlaAsnGly 635
Qy 1810 TCCTCACTAGCCCTGTACAGAGAAGTGTATGATGAGAAAGGAGAAATACCTGAAGCC 1869
Db 636 ThrProValThr-----IleAsnGlyGlnSerGlyArgLysTyrAspAla 650
Qy 1870 TGAACGGTGTGATGAGCTCTGCTGGAAGATGACGGATCCCGCTGACGTGAATGCT 1929
Db 651 TyrThrThrLeuAspIleGlyAlaLysTyrAlaValAlaGluAsnValAspLeuAsnAla 670
Qy 1930 GCGGTGAATAACCTGTCAACAGAGATTACAGTACGCTGAGCGCTGTACAGTCCCGGTAAG 1989
Db 671 AlaValTyrAsnValPheAspLysAsp----- 679
Qy 1990 AGTACGCTGTATCGCGGTATCTTCCAGACGGGATCATCAACACGGAATATGTGATA 2049
Db 680 -----ValGlyThrAspAspPheAsnThrValMet 689
Qy 2050 CCGTGGCAATTAATGATGTCGCTGAACTATCATGTC 2088
Db 690 GluGlyArgArgPheThrIleSerMetThrAlaLysPhe 702

RESULT 8
E98247
outer membrane protein irgA precursor [imported] - Agrobacterium tumefaciens (strain C58)
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C:Accession: E98247
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens
A:Reference number: A97359; PMID:21608551; PMID:11743194
A:Accession: E98247
A>Status: preliminary
A:Molecule type: DNA

```

```

A:Residues: 1-707 <KUR>
A:Cross-references: GB:AE007870; PIDN:AAK89503.1; PID:g15159377; GSPDB:GN00170
C:Genetics:
A:Gene: AGR_L1858
A:Map position: linear chromosome
C:Superfamily: ferrienterochelin receptor; tonB-dependent receptor amino-terminal homologs

Alignment Scores:
Pred. No.: 2,17e-60 Length: 707
Score: 993.00 Matches: 242
Percent Similarity: 49.80% Conservative: 133
Best Local Similarity: 32.14% Mismatches: 244
Query Match: 26.61% Indels: 134
DB: 2 Gaps: 23

US-10-625-972-4 (1-2091) x E98247 (1-707)
Qy 19 GCTTCCGTAGTCAATCCCTGTCTCGGATTTTCAGCCAGCAGCATAGCT-----GCTGCA 72
Db 26 AlaAlaMetAlaGlyThrAlaLeuGlyPheAlaValProAlaPheAlaGlnAlaSer 45
Qy 73 GAG-----GATGTGATGATGTCTCGGCATCCGCTATGAGAAAAAGCTG 117
Db 46 GluGlyAsnThrValLeuGlnGlnIleValThrAlaSerGlyPheGluGlnAsnVal 65
Qy 118 ACTAACCCAGCCGCGCAGTGTCTGTGATTAGCCAGGAGGAATTGCGATCCAGCCAGTAC 177
Db 66 LysAspAlaProAlaSerIleThrValThrArgGluAspLeuGluLysGlySerTyr 85
Qy 178 CACGATCTGGGAGGCTCTGAGATCAGTAGAGGGTGTGGATGTTGAAAGTGTGACGGT 237
Db 86 ArgAspLeuThrAspAlaLeuArgGluValGlnGlyValSerVal-----ThrGly 102
Qy 238 AAAACCCGCGGCTGGAAATCAGCATCCGAGGAATGCCAGCAGTGTACAGCTGATAC 297
Db 103 IleAlaAsnGluLysAspValPheIleArgGlyLeuProGlyAlaTyrThrLeuIleLeu 122
Qy 298 ATTGATCGTGTCTCAGCGCGGAAGCAGTGCATCCCAACCGTGTTCCTGCCATG 357
Db 123 ValAspGlyLysArgGln---SerThrArgAspAlaArgThrAsnGlyAsnSerGlyPhe 141
Qy 358 AATACCCGGTTCATGCCCTCTGCCCGCATTCAGCGTATTCAGGTTATCAGGGGGCCG 417
Db 142 GluGlnSerPheValProValSerAlaIleGluArgIleGluValValArgGlyPro 161
Qy 418 ATGTCCACACTGTATGCTCTGATGGGATGGCGGTGTGTGTAATATCATACCAAG 477
Db 162 MetSerSerLeuTyrGlySerAspAlaMetGlyValIleAsnIleThrArgLys 181
Qy 478 AATGCAGACAAATGGCTCTCTTCGTCATGAGGCTGATCTGCAGGAACACACAAA 537
Db 182 ValGlyAspValTyrSerGlySerValThrThrGluGlyThrValGlnGlnHisSerLys 201
Qy 538 TGGGGTAACAGCAGCAGTAAATTTCTGGAGCAGTGGTCCCTCTGTGATGATTCGTCTC 597
Db 202 PheGlyAsnSerGlyGlnValSerTyrTyrAlaAsnGlyProIleLeuLysAspGlnLeu 221
Qy 598 AGCCTGAGTACCGGTAGCACACACAGCGTCAAGGGTTTCATCGGTACATCATCAGGAG 657
Db 222 GlyLeuGlnLeuTyrGlyArgGlyPheThrArgGlyGluAspArgIle-----LeuAsn 239
Qy 658 GATACAGCAGCAGCGGTATCTCTTATCCACGGAGTGCACAGATTAATATCTGTGTGCA 717
Db 240 GlyThrThrGly-----AlaLysGluTyrAspPheAsnGly 251
Qy 718 CGTCTTGACTGAAGCGGTGGAGCAGGATGTCTCTGTTGATATGATACCAACCGG 777
Db 252 ArgLeuThrPheThrProAsnGluAspHisAspIleTyrLeuGlyGlyLysThrArg 271
Qy 778 CAGCGTATGATAACCGGATGGCAACTGGGAGTCTGACGGGGGATATCAGCGGACC 837
Db 272 LeuArg-----ArgAspAlaGluProGlyAspThrLeuAlaAlaThrAspAlaAsn 288

```

QY 838 CTGCGCTATAGCGAAAC---AAATTTTACGTGGCTATGATCATATCTTACCTTCGA 894
DB 289 GlyThrTyrAsnThrAsnThrArgAspHisTrpSerLeuSerHisThrGlyArgTrpGly 308
QY 895 -----ACATGGAATCGTATCTGAACCTGGAACGAG 924
DB 309 ProThrThrSerGluPheSerPheGlnGlnGluTrpAlaGluArgThrAsnPheThrArg 328
QY 925 ACAGAAAATAAAGGTCGTGAGCTTGTACGCGAGTGTACTGAAGCGCGCAAAATGGGGCTT 984
DB 329 AsnThrArgThrGlyArg-----Val 335
QY 985 GCGCGTACGCGCGGAGCTTAAGAAATCGAACCTTATCTTGAATTCATCTACTGCTTACC 1044
DB 336 ThrGluAsnProArgSerProGluLeuArgAsnThrValLeuAspGlyLysPheThrThr 355
QY 1045 CCT-----CTGGGAGATCTCATCTGTTACGGTGGGGCGAGTTTCACAGCTCGTCC 1098
DB 356 PropheGluLeuPheGlyAsnHisThrLeuValThrGlyGlyGlnTyrPheGluAlaArg 375
QY 1099 ATGAAAGACGGAGTTGTCTTCCAGCAC-----GCTGAAATCTTCCGCGCAAAAGC 1152
DB 376 LeuThrAspGlnAsnProGlyArgArgThrAspArgAspGluThrPheSerAlaThrGln 395
QY 1153 TGCTCGTATTGCTGAGATGAGTGGCATCTCACGGATGCACTTGCCTGACTCGCGGC 1212
DB 396 TrpAlaLeuPheLeuGluAspGluTrpArgIleValAspAsnPheAlaLeuThrGlyGly 415
QY 1213 AGCCCTATGACATCATGACAAATCGGGGACACTTCAGTCCGCGTGCATATCTGTC 1272
DB 416 LeuArgLeuAspAsnHisGluLysTyrGlyAsnHisPheSerProArgLeuTyrGlyVal 435
QY 1273 TGGGATGTCGAGATCCCTGAGCGTGAAGCGGTGACCCAGGATATAAGCCACC 1332
DB 436 TrpSerAlaThrGluGluLeuThrIleLysGlyValSerThrGlyPheArgAlaPro 455
QY 1333 AGAATGGCGACGACATATAAAGCGATTAGTGTGTCTCGGCGAGGAAAAACAAT--- 1389
DB 456 GluIleArgGluIleAlaProGlyTyrAlaTyrThrGlyGlyGlyCysSerTyr 475
QY 1390 -----CTACTGTGTAACCCGACCTGAAGCGGAGAGAGAGC 1425
DB 476 GlyProSerGlyThrCysGlyValIleGlyAspProAsnLeuGluAlaGluLysSer 495
QY 1426 CTCAGTTATGAGCTGGGTGTATACGATACCCCGCGTCTCAATGCCAATGTACA 1485
DB 496 ThrSerTyrGluIleAlaAlaLeuThrAspAsn---GlyAspIleAlaLeuGlyAlaThr 514
QY 1486 GGTTTTATGACTCTTCTCCAAAGATT----- 1515
DB 515 TyrPheTyrThrAspPheLysAspLysIleSerAsnAlaLeuValLeuAsnProAspGly 534
QY 1516 -----GTCTCTATTCCATAAT 1533
DB 535 ThrProAlaArgTrpSerGluAspArgAsnTyrArgLeuTrpTyrAsnTyrAsnIleAsp 554
QY 1534 GATAACACCAATAGCTATGTAAACAGCGGAGGCGGCTTGCACGGTGTGAAATTCGC 1593
DB 555 Asp-----AlaIleIleGlnGlyValGluLeuThr 564
QY 1594 GGCACATTCCCGTGTGG-----TCAGAGGATGTACCGCTGTCTACTGAATTACACTGG 1647
DB 565 AlaThr-----TrpTyrAlaThrSerGluLeuThrLeuArgGlyAsnTyrThrTyr 581
QY 1648 ACCGGAAGTGAACAACGTGTGTGATACAAAGGTGCGCGCTGAGTTATACCCCTGAA 1707
DB 582 ThrHisSerGluGlnLysThrGlyAspTyrGluGlyPheProLeuAlaArgThrProGlu 601
QY 1708 CACATGGTGAATCGGAAACTGACCTGCGAGATCAC-----GAAGAGGTGCATCATGG 1761
DB 602 HisMetAlaAsnLeuArgGlyAspTrp---ValThrProIleAspGlyLeuGluAlaTrp 620
QY 1762 CTGGGTGCCGTTATCGCGGGAACA-----CCAGTTTCCACCAAGATTAT 1809

DB 621 ValSerLeuAsnTyrHisGlySerGluIleAsnAlaGlyProArgIleGlyAlaAsnGly 640
QY 1810 TCCTACTACTGACCGCTCTACAGAAAGAGTGTATGATGAGAAAGGAGATATCCTGAAAGCC 1869
DB 641 ThrProValThr-----IleAsnGlyGlnSerGlyArgLysTyrAspAla 655
QY 1870 TGGACCGTGTGATCGACGTCTGCTGGAAGATCAACGATCCCTGACGCTGAATGCT 1929
DB 656 TyrThrThrLeuAspIleGlyAlaLysTyrAlaValAlaGluAsnValAspLeuAsnAla 675
QY 1930 GCGGTCAATAACCTGCTCAACAGGATTACAGTACGCTGAGCCTGTACAGTCCGGTAAAG 1989
DB 676 AlaValTyrAsnValPheAspLysAsp----- 684
QY 1990 AGTACGCTGTATCCCGGTGATCTACTCCAGCGGATCATCAACACAGATATGTGATA 2049
DB 685 -----ValGlyThrAspAspPheAsnThrValMet 694
QY 2050 CCTGACGGAATTACTGGATGTCTGCTGAATCATCTACGTTTC 2088
DB 695 GluGlyArgArgPheIleSerMetThrAlaLysPhe 707
RESULT 9
AG0782
collicin I receptor precursor [imported] - Salmonella enterica subsp. enterica serovar Typhimurium
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AG0782
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Garra, P.
Nature 413, 848-852, 2001
A:Authors: Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Garra, P.
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhimurium
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AG0782
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-663 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD02581.1; PID:G16503438; GSPDB:GN00176
C:Genetics:
A:Gene: STY2434
C:Superfamily: ferrienterochelin receptor; tonB-dependent receptor amino-terminal homolo

Alignment Scores:
Pred. No.: 5.59e-50 Length: 663
Score: 841.50 Matches: 235
Percent Similarity: 48.03% Conservative: 107
Best Local Similarity: 33.01% Mismatches: 263
Query Match: 22.55% Indels: 107
DB: 2 Gaps: 21

US-10-625-972-4 (1-2091) x AG0782 (1-663)

QY 61 ATAGCTCTCGAGAGGAT-----GTGATGATGTCCTCGGATCCGCTATGAGAAA 111
DB 23 ValAlaAlaAlaThrAspAspGlyGluThrMetValValThrAlaSerAlaIleGluGln 42
QY 112 AAGTGTACTAACGACGCGCGGAGGTTCTGTGATTAGCCAGGAGGATTCAGTCCAGC 171
DB 43 AsnLeuLysAspAlaProAlaSerIleSerValIleThrGlnGlnAspLeuGlnArgArg 62
QY 172 CAGTACCACCATCTGGCGGAGGCTCTGAGATCATAGAGGGTGTGGATGTT---GAAAGT 228
DB 63 ProValGlnAsnLeuLysAspValLeuLysGluValProGlyValGlnLeuThrAsnGlu 82
QY 229 GGTACGGGTAACCCGAGGCGGCTGGAATCAGCATCCGAGGATCCAGCCAGTTACACG 288
DB 83 GlyAspAsnArgLysGly-----ValSerIleArgGlyLeuAspSerSerTyrThr 99
QY 289 CTGATACTGATTGATGGTGTTCGTGAGGCGGAGAGCAGTACGCTGACTCCCAACGGTTT 348

Db 599 SerAlaThrAlaLysThrProGlyGlyTyrThrIleTrpAsnThrGlyAlaAlaTrpGln 618
 Qy 1903 ATGACGATGCGCTGACGCTGAATGCTGCGGTGAATAACCTGCTCAACAGGATTACAGT 1962
 Db 619 ValThrLysAspValLysLeuArgAlaGlyValLeuAsnLeuGlyAspLysLeuSer 638
 Qy 1963 GACGTGACCTGTACAGTCCGCTAGAGTACGCTGTATGCCGGTGATTACTTCCAGAGC 2022
 Db 639 -----ArgAspAspTyr----- 642
 Qy 2023 GGATCATCAACACAGGATATGTGATACCTGACGAGAAATPACTGGATGCTGCTGACTAT 2082
 Db 643 -----SerTyrAsnGluAspGlyArgGlyPheMetAlaValAspTyr 657
 Qy 2083 CAGTTC 2088
 Db 658 ArgPhe 659
 RESULT 11
 QREIC
 colicin I receptor precursor - Escherichia coli (strain K-12)
 C:Species: Escherichia coli
 C>Date: 30-Sep-1990 #sequence revision 05-Dec-1997 #text change 01-Mar-2002
 C:Accession: B64984; A32056; A33868; A28377; C41871; A35408; S24561
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, D.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:97426617; PMID:9278503
 A:Accession: B64984
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-663 <BLAT>
 A:Cross-references: GB:AE000304; GB:U00096; NID:G1788470; PID:NAC75216.1; PID:G1788478;
 A:Experimental source: strain K-12, substrain MG1655
 R:Nau, C.D.; Konisky, J.
 J. Bacteriol. 171, 1041-1047, 1989
 A:Title: Evolutionary relationship between the TonB-dependent outer membrane transport protein and the colicin I receptor.
 A:Reference number: A32056; MUID:89123100; PMID:2644220
 A:Accession: A32056
 A:Molecule type: DNA
 A:Residues: 1-146; RCARCSYHQ, 159, 'NRSEMV', 166, 'YRYRHHSGTSRSR', 182-527, 'N', 529-613,
 A:Note: the authors translated the codon AAT for residue 528 as Ile
 R:Nau, C.D.; Konisky, J.
 J. Bacteriol. 171, 4530, 1989
 A:Reference number: A33868
 A:Contents: corrections
 A:Accession: A33868
 A:Molecule type: DNA
 A:Residues: 144-184; 610-663 <NA2>
 R:Griggs, D.W.; Konisky, J.
 J. Bacteriol. 169, 5343-5352, 1987
 A:Title: Cloning and promoter identification of the iron-regulated *cir* gene of Escherichia coli.
 A:Reference number: A28377; MUID:88058737; PMID:3316180
 A:Accession: A28377
 A:Molecule type: DNA
 A:Residues: 1-59 <GRI>
 A:Cross-references: GB:M1295; NID:G145545; PID:NAA23581.1; PID:G551794
 A:Note: residues 26-45 were confirmed by protein sequencing
 R:Steffes, C.; Ellis, J.; Wu, J.; Rosen, B.P.
 J. Bacteriol. 174, 3242-3249, 1992
 A:Title: The *lysP* gene encodes the lysine-specific permease.
 A:Reference number: A41871; MUID:92250419; PMID:1315732
 A:Accession: A41871
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-96, 'D', 98-125 <STE>
 A:Cross-references: GB:M89774; NID:9466776; PID:NAA17054.1; PID:G466779; EMBL:X65029
 R:Griggs, D.W.; Kafka, K.; Nau, C.D.; Konisky, J.
 J. Bacteriol. 172, 3529-3533, 1990
 A:Title: Activation of expression of the Escherichia coli *cir* gene by an iron-independent promoter.
 A:Reference number: A35408; MUID:90264362; PMID:2160948
 A:Accession: A35408

A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-21 <GR2>
 C:Comment: This outer membrane receptor for colicins Ia and Ib is regulated by both cell membrane transport proteins.
 C:Genetics:
 A:Gene: *cir*; *cirA*; *feuA*
 A:Map position: 43 min
 C:Keywords: iron transport; membrane protein
 C:Keywords: iron transport; membrane protein
 F:1-25/Domain: signal sequence #status predicted <SIG>
 F:26-624/Product: colicin I receptor #status experimental <MAT>
 F:66-211/Domain: tonB-dependent receptor amino-terminal homology <TNC>
 F:363-663/Domain: tonB-dependent receptor carboxyl-terminal homology <TNC>
 Alignment Scores:
 Pred. No.: 2,33e-49 Length: 663
 Score: 832.50 Matches: 234
 Percent Similarity: 47.45% Conservative: 120
 Best Local Similarity: 31.37% Mismatches: 237
 Query Match: 22.31% Indels: 155
 DB: 1 Gaps: 25
 US-10-625-972-4 (1-2091) x QREIC (1-663)
 Qy 40 CTCGGATTTTCAGCCAGCAGCATAGCTGCTGCA----- 72
 Db 10 ValGlyLeuCysLeuSerAlaIleSerCysAlaTrpProValLeuAlaValAspAsp 29
 Qy 73 GAGGATGTGATGTTCTCGGATCCCGGCTATGAGAAAGCTGACTAACGAGCGGCC 132
 Db 30 GlyGluThrMetValValThrAlaSerSerValGluGlnAsnLeuLysAspAlaProAla 49
 Qy 133 AGTGTCTGTGATTACCCAGGAGAAATTCAGTCCAGCCAGTACCGATCTGGCGGAG 192
 Db 50 SerIleSerValIleThrGlnGluAspLeuGlnArgLysProValGlnAsnLeuLysAsp 69
 Qy 193 GCTCTGAGATCAGTAGAGGCTGTGGATGTT---GAAAGTGTGTACGGTAAACCGGAGG 249
 Db 70 ValLeuLysGluValProGlyValGlnLeuThrAsnGluGlyAspAsnArgLysGly--- 88
 Qy 250 CTGAAATCAGCATCCGAGAAATCCGACGATTCACGCTGATACGTATGATGTGTT 309
 Db 89 -----ValSerIleArgGlyLeuAspSerSerTyrThrLeuIleLeuValAspGlyLys 106
 Qy 310 CGTCAGGCGGAGGAGTACGAGTACCTCCCAACGGTTTTTCTGCCATGAATACCGGGTTC 369
 Db 107 ArgValAsnSerArgAsnAlaValPheArgHisAsnAspPheAspLeuAsn----- 123
 Qy 370 ATGCCCCCTCTGCGCCGCTATGAGCGTATTGAGTTATCAGGGGCGCGATGTCCACTG 429
 Db 124 TrpIleProValAspSerIleGluArgIleGluValValArgGlyProMetSerSerLeu 143
 Qy 430 TATGGCTCTCATGCGATGGCGGCTGTGGTGAATATCATTACCAAGAAAGATGCAGACAA 489
 Db 144 TyrGlySerAspAlaLeuGlyGlyValValAsnIleIleThrLysLysIleGlyGlnLys 163
 Qy 490 TGGCTCTCTTCCTCAATGTCAGGCGCTCAATCTCCAGGAAAGCAACAAATGGGTAAACAGC 549
 Db 164 TrpSerGlyThrValThrValAlaAspThrThrIleGlnGluHisArgAspArgGlyAspThr 183
 Qy 550 AGCCAGTTTAATTTCTGGAGCAGTGGTCCCTTGTGGATGATCTCTCAGCCCTGCGAGTGA 609
 Db 184 TyrAsnGlyGlnPhePheThrSerGlyProLeuIleAspGlyValLeuGlyMetLysAla 203
 Qy 610 CGCGGTAGCACACACACAGCGT-----CAGGGTTTCATCGCTCACTCACCTG 654
 Db 204 TyrGlySerLeuAlaLysArgGluLysAspAspProGlnAsnSerThrThrThrAspThr 223
 Qy 655 AGCGATACAGGCGGCGGCTATCTCTTATCCACCGAGTACAGAAATTATTAATCTTGGT 714
 Db 224 GlyGluThr----- 226

[illegible]

QY	1207	CGCGGAGCCGCTATGAACATCATATGACCAANTCGGGGACACTTCAGTCGCGCTGCATAT	1267
Db	380	ThrGlyValArgMetAspAspHisGluThrTyrGlyGluHisTrpSerProArgAlaTyr	399
QY	1267	CTGTCTCGGATGTGGCAGATGCTGTGACGCTGAAGCCGCTGCACACGGATATAAG	1326
Db	400	LeuValTyrAsnAlaThrAspThrValThrValValGlyGlyGlyTrpAlaThrAlaPheLys	419
QY	1327	GCACCCAGATGGGCGACCTACATAAAGGGATTATGTGTGTCTCCGGCGACGGAAAAACA	1386
Db	420	Ala***SerLeuLeuGlnLeuSerProAspTrpThrSerAsnSerCysArgGlyAlaCys	439
QY	1387	AATCTACTGTGTAAACCCGACCTGAAGCCGGAAGAGCGCTCAGTTATGAGGCTGGGGT	1446
Db	440	LysIleValGlySerProAspLeuLysProGluThrSerGluSerTrpGluLeuGlyLeu	459
QY	1447	TATTACGATAACCCCC-----GCCGCTCTGAATCCCAATGTCACAGGTTTATG	1494
Db	460	TyrTrpMetGlyGluGluGlyTrpLeuGluGlyValGluSerSerValThrValPheArg	479
QY	1495	ACTGACTCTCCAACAGATTGCTCTTATTCATAAATGATAACCACAAT-----	1545
Db	480	AsnAspValLysAspArgIle---SerIleSerArgThrSerAspValAlaAlaPro	498
QY	1546	-----ACGTATGTAAAC	1557
Db	499	GlyTyrGlnAsnPheValGlyPheThrAlaAsnGlyValProValPheSerTyrTyrAsn	518
QY	1558	AGCGAAAGCCCGGTTCACGGTGTGGAAATTTGCGGCACATGTGCGCTGTGGTCAGAG	1617
Db	519	ValAsnLysAlaArgIleGlnGlyValGluThrGluLeuLysIlePro---PheAsnAsp	537
QY	1618	GATGTACGCTGTCTACGTAAATTCACCTGACCCGAGTGAACAAAGTGAT-----	1668
Db	538	GluTrpLysLeuSerLeuAsnTyrThrTyr-----AsnAspGlyArgAspValSerAsn	555
QY	1669	GGTGATAACAAAGGTGCGCGCTGAGTTATACCCCTGAACACATGTGTGAATGCGAAACTG	1728
Db	556	GlyGluAsnLys-----ProLeuSerAspLeuProPheHisThrAlaAsnGlyThrLeu	573
QY	1729	AACGTGGCAG-----ATCACCGAAGAGTGGCATCATGGCTGGGTGCCGCTTATCCGGG	1782
Db	574	AspTrpLysProLeuAlaLeuGluAspTrpSerPheTyrValSerGlyHisTyrThrGly	593
QY	1783	AAAAACACCGCTTTCACCCAGAAATTATTGTCACCTGAGCGCTGTACAGAAAGATGTAT	1842
Db	594	-----GlnLysArgAlaAsp	598
QY	1843	GATGAGAAAGGAGAAATCACTGAAGCCTGGACGGTGGTGGATCGAGTCTGCTCGTGAAG	1902
Db	599	SerAlaThrAlaLysThrProGlyTyrThrIleTrpAsnThrGlyAlaAlaTrpGln	618
QY	1903	ATCACGATGCCCTGACGCTGAATGCTGCGGTGAATAACCTGCTCAACAGGATTACAGT	1962
Db	619	ValThrLysAspValLysLeuArgAlaGlyValLeuAsnLeuGlyAspLysAspLeuSer	638
QY	1963	GAGCTGAGCCTGTACGTATGCGCGGTAAAGTACCTGTATGCCGGTACTTCTCCAGACG	2022
Db	639	-----ArgAspAspTyr-----	642
QY	2023	GGATCATCAACACAGGATATGTGATACCTGACGCGMAATTACTGGATGTCTGCTGAACAT	2082
Db	643	-----SerTyrAsnGluAspGlyArgTyrPheMetAlaValAspTyr	657
QY	2083	CAGTTC	2088
Db	658	ArgPhe	659

RESULT 13

probable TonB-dependent receptor PA1922 [imported] - *Pseudomonas aeruginosa* (strain PA01)
 F83405
 C:Species: *Pseudomonas aeruginosa*
 C:Date: 15-Sep-2000 #sequence revision 15-Sep-2000 #text change 11-Dec-2000

C;Accession: F83405
R;Sover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic patho
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: F83405
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-653 <STO>
A;Cross-references: GB:AE004618; GB:AE004091; NID:g9947912; PIDN:AA05310.1; GSPDB:GN001
A;Experimental source: strain PA01
C;Genetics:
A;Gene: PA1922
C;Superfamily: ferrienterochelin receptor; tonB-dependent receptor amino-terminal homolo

Alignment Scores:	
Pred. No.:	7.28e-43
Score:	738.00
Pred. Similarity:	48.60%
Percent Similarity:	40.93%
Best Local Similarity:	30.93%
Query Match:	19.77%
DB:	2
Length:	653
Matches:	210
Conservative:	120
Mismatches:	245
Indels:	104
Gaps:	23

US-10-625-972-4 (1-2091) x F83405 (1-653)

Qy	28	GTGATTCCTGCTCTCGGATTTTCAGCCAGCAGCATAGCTGCTGCAGAGATGTG-----	81
Db	24	LeuLeuProCysLeu-----AlaLeuAlaProValAspLeuGluPro	38
Qy	82	ATGATTCTCGGCATCCGGCTATGAGAAAAGTGACTAACCGCAGCGCCAGTGTTTCT	141
Db	39	ThrValValSerAlaThrThrThrGluArgLysLeuArgAspAlaProAlaSerValSer	58
Qy	142	GTGATTAGCAGGAGGAATTCAGTCCAGCCAGTACCAGCATCTGGCGGAGGCTCTGAGA	201
Db	59	LeuIleAlaAlaGluAspLeuArgArgProValArgAspLeuGluAlaLeuArg	78
Qy	202	TCAGTAGAGGTGTGGATGTTTGAAGT---GGTACGGGTAAAAACCGGAGGCTGGAATC	258
Db	79	GlySerGluSerLeuGlnPheAsnGlyValGlyMetSerArgArgGly-----Ile	95
Qy	259	AGCATCCGAGGAATCCAGCCAGCATACCGTGTATCTGTTGATGTGTTCGTGAGGC	318
Db	96	SerValArgGlyMetSerSerGluHisThrLeuValLeuValAspGlyGlnArgLysSer	115
Qy	319	GCAAGCAGTGCAGTGACTCCCAACGGTTTTTCTGCCATGAATACCGGGTTCATGCCCCCT	378
Db	116	ThrSerSerGlyAlaIleAlaHis-----SerAspPheAspLeuGlyTrpVal---Pro	132
Qy	379	CTGGCCGCATTGACGTATTGAGGTATCAGGGGCGGATGTCACACATCTATGGCTCT	438
Db	133	ValGluSerIleGluArgIleGluValValArgGlyProMetSerSerLeuArgGlySer	152
Qy	439	GATGGATGGCGGTGTGGTGAATATCATTAACGAGAAAGTAATCGAGCAATGGCTCTCT	498
Db	153	GluAlaLeuGlyValValAsnValIleThrArgAlaThrAspThrTrpThrGly	172
Qy	499	TCGGTCAATGCAGGCTGAATCTGCAGAAAGCAACAATGGGGTAACAGCAGCAGTTT	558
Db	173	SerGlyLeuLeuAspGlyGlyValValArgGluAspGlyLeuGlyGlyLysIleGlnLeu	192
Qy	559	AATTTCTGGAGCAGTGTGCCCTTGTGGATGATTCGTACGCCCTGCAGGTACGGGTAGC	618
Db	193	GlyAlaTyrLeuAlaGlyProLeuValProGlyLysLeuGlyLeuAlaLeuAsnGlyGlu	212
Qy	619	ACACAACAGCTCAGGGTTCA-----TCGGTCACTCATCTG-----	654
Db	213	SerArgArgGlnGlnGluThrProAspAlaAspGluArgArgLeuSerGluLeuGluGly	232
Qy	655	-----AGCGATACAGCAGGACGGCTATTCCTTAT---CCACGGAGTCAAGAAATTAT	705

233	GlyAsnAlaAspSerGlyGlyLeuAsnLeuSerTrpThrProAspAspAlaGlnArgIle	252
706	AATCTTGGTGCAGCTCTTGAC-----TGAAGCGCGTCGAGCAGGATGTGCTC	753
253	AspLeuGlyHisGlnArgGlyArgGluArgTrpArgAsnSerGlu-----	268
754	TGTTTGAATGATGATACCAACCGGACGCTTATGATAACCGGATGGGCAACTGGGGACT	813
268	-----	268
814	CTGACGGGGGA-----TATGACCGGACCGCTCGCTATGAGCGAAACAAA	858
269	---ThrGlyGlyProArgSerArgTyrTyrGluSerArgAspValIleGluArgGluArg	287
859	ATTTCAGCTGGCTATGATCATACTTTCACCTTCGGAACATGGAATTCGTATCTGAAGCTG	918
288	TrpSerLeuAlaHisAsn-----GlyGlnTrp-----AspTrp	298
919	AACGAGACAAATAAAGTCTGTGAGCTGTACGCAGTGTACTGAAGCGCGCAACAATGG	978
299	GlySerSerGlnLeuArg-----ThrTyrArgAsnArgLeuGluArgHisAsnAla	315
979	GGGCTTGCCGGTCAG-----CGCGGGAGCTTAAAGAAATCGAACCTTATCTCTG	1026
316	ArgSerAspGlyGlnProProSerAsnProGlnArgLeuThrAspSer-----ValVal	333
1027	AATTCATTACTCTTACCTCTCTGGGAGATCTCATCTGTTACGCTGGGGCGCAGTTT	1086
334	AspGlyHisLeuSerValProAlaPheGluArgHisLeuPheThrLeuGlyGlyGluTrp	353
1087	CAGAGCTCGTCCATGAAGACGGAGTTGTCCTTGCACGACGACAGTGAAACTTCCGGCAG	1146
354	ArgLysGluGluLeuGluAspArgSerValAsnThrAlaGlyAspAlaSerAlaArgHis	373
1147	AAAAGCTGTGCGTATTTGCTGAGGATGAGTGCATCTCACGGATGCACATTGCGCTGACT	1206
374	Lys-----AlaLeuPheLeuGlnAspGluIleAlaPheGlyProAspTrpSerLeuThr	391
1207	CGGGCAGCGCTATGAACATCATGACCAATTCGGGGACACTTCATCTCCGCTGCATAT	1266
392	LeuGlySerArgPheAspLysHisGluAlaPheGlyTrpGluSerSerProArgLeuTyr	411
1267	CTGTCTGGATGTGGCAGATGCGCTGCAGCTGAAAGCGGTGTGACCGGATATAAG	1326
412	LeuLeuHisHisLeuSerAspValLeuThrLeuArgAlaGlyValGlyArgGlyTyrLys	431
1327	GCACCCAGATGGGCGAGCTACATAAAGGATTAGTGGTGTGTCGGGCGAGGAAACA	1386
432	AlaProSerLeuLysGlnLeuSerProGluTyrAlaAlaValGlyGlyGlyArgPhe	451
1387	AATCTACTTGTAAACCCCGACCTGAAACCGGAGAGAGCGTCTAGTTATGAGGCTGGGGT	1446
452	ThrIleTyrGlyAsnProAspLeuLysProGluThrAsnThrSerTyrGluLeuGlyAla	471
1447	TATTACGATAACCCCGCGCTCGAATGCCAATGTCACAGGTTTATGACTGACTTCTCC	1506
472	AspTyrGlnGlyAsp---GlyTrpSerLeuLysGlyThrValPheGluAsnAspValArg	490
1507	AACAAGATTGTCTCTATTTCATA-----AATGATAACACCAATPAGC	1548
491	AspLeuIleGlnThrValCysValAlaArgCysGlyValArgGlyGlyGluIleArgAsn	510
1549	TATGTAACACCGGAAGGCCGCTTGCACGGTCTGGAATTCGCGGCACATTCGCGCTG	1608
511	TyrGluAsnValAspArgAlaArgIleArgGlyLeuGluLeuGlyGlyValAspLeu	530
1609	TGGTCAGAGGATGTCAAGCTGTCACTGAATTACACCTGGACCCCGAAGTGAACACACGTGAT	1668
531	---ProAlaAspLeuArgTrpGluLeuAsnTyrThrTyrLeuAspAlaArgAsnArg---	548
1669	GGTGATAACAAGGTGGCCGCTGAGTTATACCCCTGAACACATGTGTGATGCGAACTG	1728
549	-----ThrAlaGlyGlnArgLeuGlyAspArgSerArgHisLeuAlaAsnSerGlnLeu	566

Db 488 TyrLysAlaProSerLeuTyrGlnThrAsnProAsnTyrIleLeuTyrSerLysGlyGln 507
Qy 1360 AGTGGTGTCCGGGAGGAAACAAATCTACTTGGTAACCCGACCTGAGCGGAA 1419
Db 508 GlyCysTyrAlaSerAlaGlyGlyCysTyrLeuGlnGlyAsnAspAspLeuAlaGly 527
Qy 1420 GAGAGCGTCACTATGAGCGTGGGGTGTATTACGATAACCCCGCGTCTGAATGCCAAT 1479
Db 528 ThrSerIleAsnLysGluIleGlyLeuGluPheLysArgAspGlyTrpLeu--AlaGly 546
Qy 1480 GTCACAGGTTTATGACTGCTCTCCAAAGATGTCTCT-- 1521
Db 547 IleThrTrpPheArgAsnAspTyrArgAsnLysIleGluAlaGlyTyrValAlaValGly 566
Qy 1522 ---TATTCATTAATATGATACCAATAGCTATTAACAGCGGAAAGCCCGGTTCAC 1578
Db 567 GlnAsnAlaValGlyThrAspLeuTyrGlnTrpAspAsnValProLysAlaValValGlu 586
Qy 1579 GGTGTGAATTTGCCGACATTCGCG-- 1623
Db 587 GlyLeuGluGlySerLeuAsnValProValSerGluThrValMetTrpThrAsnAsnIle 606
Qy 1624 ACGTGTCACTGAATTACACCTGGACCCGGAAGTGAACAAACGTCATGATTAACAAAGT 1683
Db 607 ThrTyrMetLeu-----LysSerGluAsnLys-----ThrThrGly 618
Qy 1684 CGCGCGTGTAGTTATACCTTGACACATGGTGTGAATGCGAACTGAACCTGCGAGATCAC 1743
Db 619 AspArgLeuSerIleIleProGluTyrThrLeuAsnSerThrLeuSerTrpGlnAlaArg 638
Qy 1744 GAAGAGTGGCATCATGGCTGGGTGGCGGTTATCGCGGGGAAACACACACCGCTTCCACCCAG 1803
Db 639 GluAspLeuSerMetGlnThrThrPheThrTrpTyrGlyLysGlnLysProLysLysTyr 658
Qy 1804 AATTATTCGTCACTGAGCGCTGTACAGAAAGTGTATGATGAGAAAGAGATACCTG 1863
Db 659 AsnTyrLysGlyGlnProAlaValGlyProGluThrLysGlu-----Ile 673
Qy 1864 AAAGCCTGGAGCGGTGGTGGATGCGGTCTGTCTGGAAGATGACGATGCCCTCGCGCTG 1923
Db 674 SerProTyrSerIleValGlyLeuSerAlaThrTrpAspValThrLysAsnValSerLeu 693
Qy 1924 AATGCTCGGTGAATTAACCTGCTCAACAGAGTATCAGTGCAGCTGAGCGCTGACAGTCC 1983
Db 694 ThrGlyGlyValAspAsnLeuPheAspLys-----ArgLeuTrpArgAla 708
Qy 1984 GGTAAAGATACGCTG-----TATGCGGTGATTACTTCAGACGCGGATCATCAACA 2034
Db 709 GlyAsnAlaGlnThrThrGlyAspLeuAlaGlyAlaAsnTyrIleAlaGlyAlaGlyAla 728
Qy 2035 ACAGGATATGATGATCTACGAGGAAATTAATCTGATGCTGCTGATCACTATCAGTTC 2088
Db 729 TyrThrTyrAsnGluProGlyArgThrTrpTyrMetSerValAsnThrHisPhe 746
RESULT 15
B85557
hypothetical protein fepA [imported] - Escherichia coli (strain O157:H7, substrain EDL93
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: B85557
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamocis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; PMID:21074935; PMID:11206551
A;Accession: B85557
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-746 <STO>
A;Cross-references: GB:AE005174; NID:gl2513473; PIDN:AGS4918.1; GSPDB:GN00145; UWGP:Z07
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: fepA

C;Superfamily: ferrienterochelin receptor; tonB-dependent receptor amino-terminal homolog

Alignment Scores:
Pred. NO.: 2.33e-35 Length: 746
Score: 629.00 Matches: 210
Percent Similarity: 45.25% Conservative: 133
Best Local Similarity: 27.70% Mismatches: 285
Query Match: 16.85% Indels: 130
DB: 2 Gaps: 29

US-10-625-972-4 (1-2091) x B85557 (1-746)

Qy 73 GAGGATGTGATGTTCTCTGGCATCCGGCTATGAGAAAAGCTGACTAACGACGCGGC 132
Db 33 AspAspThrIleValValThrAlaAla-----GluGlnAsnLeu-----GlnAlaPro 48
Qy 133 AGTGTCTCTGATTACCCAGGAGGATTCAGTCCAGCCAGCTAC---CACATCTGGCG 189
Db 49 GlyValSerThrIleThrAlaAspGluIleArgLysAsnProValAlaArgValSer 68
Qy 190 GAGGCTCTGATCAGTACGAGGCTGTGGATTT-----GAAAGTGTACGGGTAAAACC 243
Db 69 GluIleIleArgThrMetProGlyValAsnLeuThrGlyAsnSerThrSerGlyGlnArg 88
Qy 244 GGA---GGGTGGAATCAGCATCCGAGGAATGCCAGCAGTACACCGCTGATCTGATT 300
Db 89 GlyAsnAsnArgGlnIleAspIleArgGlyMetGlyProGluAsnThrLeuIleLeu 108
Qy 301 GATGGT-----GTTCTGTCAGGGCGGAAGCAGTACGCTGACT 336
Db 109 AspGlyLysProValSerSerArgAsnSerValArgGlnGlyTrpArgGlyGluArgAsp 128
Qy 337 CCACAGGTTTCTGCCATGATACCGGTTTCATGCCCTCTGCGCCGCTGAGCGT 396
Db 129 ThrArgGly-----AspThrSerTrpValProPro---GluMetIleGluArg 143
Qy 397 ATTGAGGTTATCAGGGGCGGATGTCCACATGTATGCTCTGTATGCTGATGCGGTGGCTGTG 456
Db 144 IleGluValLeuArgGlyProAlaAlaAlaArgTyrGlyAsnGlyAlaAlaGlyVal 163
Qy 457 GTCAATATCATTCACGAGAAGAATGACAGCAATGGCTCTCTCCGTCAATGACGGGCTG 516
Db 164 ValAsnIleIleThrLysLysGlySerGlyGluTrpHisGlySerTrpAspAlaTyrPhe 183
Qy 517 AATCTGCAGGAAGCAACAAATGGGTAAACAGCAGCCAGTTTAAATTCTTGGAGCAGTGGT 576
Db 184 AsnAlaProGluHisLysGluGluGlyAlaThrLysArgThrAsnPheSerLeuThrGly 203
Qy 577 CCCCTGTGATGATCTCTGACGCTGACGCTGACGCTGACGCTGACGCTGACGCTGACGCT 636
Db 204 ProLeu---GlyAspGluPheSerPheArgLeuTyrGlyAsnLeuAspLysThrGlnAla 222
Qy 637 TCATCG-----GTCACATCACTGACGCGGATGACAGCAGCGGCGGT 675
Db 223 AspAlaTrpAspIleAsnGlnGlyHisGlnSerAlaArgAlaGlyThrTyrAlaThrThr 242
Qy 676 ATTCTTTATCCACGAGTACAG---AATTATATCTTGTGTCACGCTTGTGACGAGG 732
Db 243 LeuProAlaGlyArgGluGlyValIleAsnLysAspIleAsnGlyValValArgTrpAsp 262
Qy 733 GCGTCGAGCAGGATGCTGCTGTTGATATGATATGATATGATATGATATGATATGATATGAT 780
Db 263 PheAlaProLeuGlnSerLeuGluLeuGluAlaGlyTyrSerArgGlnGlyAsnLeuTyr 282
Qy 781 -----CGTTATGATTAACCGGATGGCAACTGGGGAGTCTGACGGGGGATAT 828
Db 283 AlaGlyAspThrGlnAsnThrAsnSerAspAlaTyrThrArgSerLysTyrGly----- 300
Qy 829 GACCGGACCTGCGCTATCAGCGAAACAAATTTCA-----GCTGGCTATGAT 876
Db 301 AspGluThrAsnArgLeuTyrArgGlnAsnTyrSerLeuThrTrpAsnGlyGlyTrpAsp 320
Qy 877 CATCTTTTACCTTCGGAAACATCGTAATCGTATCTGAACTCGAAGACGACAGCAATAA 936


```

Db      321  AsnGlyValThrThrSerAsnTrpValGlnTyr-----GluHisThrArgAsn--- 336
QY      937  GGTGGTGGAGCTGTACGCGAGTGTACTGAAGCGCGACAAATGGGGGCTTGCCTG--- 990
Db      337  -----SerArgIleProGluGlyLeuAlaGlyGlyThr 347
QY      991  -----CAGCGCGGAGCTTAAGAAATCGAACCTT----- 1020
Db      348  GluGlyLysPheAsnGluLysAlaThrGlnAspPheValAspAsnAspLeuAspVal 367
QY      1021  ATCTCGAATTCATTACTGCTTACCCCTCTCGGA-----GAATCTCATCTGGTTAGCGTG 1074
Db      368  MetLeuHisSerGluValAsnLeuProileAspPheLeuValAsnGlnThrLeuThrLeu 387
QY      1075  GGGCGCGAGTTTCAGACTCGTCCATGAAA----- 1104
Db      388  GlyThrGluTrpAsnGlnGlnArgMetLysAspLeuSerAsnThrGlnAlaLeuThr 407
QY      1105  -----GACGGAGTTGCTCTTCCACGACAGTGAACCTTTC 1140
Db      408  GlyThrAsnThrGlyAlaIleAspGlyValSerAlaThrAspArgSerProTyrSer 427
QY      1141  CGGCAGAAAAGCTGGTGGTATTGCTGAGGATGAGTGGCATCTCACGGATGCGTGGC 1200
Db      428  LysAlaGluIlePheSerLeuPheAlaGluAsnMetGluLeuThrAspSerThrIle 447
QY      1201  CTGACTCGGGGCGAGCCCTATGACATCATGAGCAATTCGGGGGACACTTCAGTCCGCGT 1260
Db      448  ValThrProGlyLeuArgPheAspHisHisSerIleValGlyAsnAsnTrpSerProAla 467
QY      1261  GCATATCTGCTGGGATGGCAGATGCTGACGCTGAAGCGGCTGTGACCGCGGA 1320
Db      468  LeuAsnIleSerGlnGlyLeuGlyAspPheThrLeuLysMetGlyIleAlaArgAla 487
QY      1321  TATAAGCACCCAGATGGGGCAGCTACAT-----AAAGGATT 1359
Db      488  TyrLysAlaProSerLeuTyrGlnThrAsnProAsnTyrIleLeuTyrSerLysGlyGln 507
QY      1360  AGTGGTGTGTCGGCGAGGAAAACAAATCTACTTGTAAACCCCGCTGAGCGCGNA 1419
Db      508  GlyCysTyrAlaSerAlaGlyGlyCysTyrLeuGlnGlyAsnAspLeuLysAlaGlu 527
QY      1420  GAGAGCGTCAATTATGAGGTGGGTGTATTACGATAACCCCGCTGTAATGCCAAT 1479
Db      528  ThrSerIleAsnLysGluIleGlyLeuGluPheLysArgAspGlyTrpLeu--AlaGly 546
QY      1480  GTCACAGGTTTATGACTGACTTCTCCACAGATTGCTCT----- 1521
Db      547  IleThrTrpPheArgAsnAspTyrArgAsnLysIleGluAlaGlyTyrValAlaValGly 566
QY      1522  ---TATTTCATAAATGATAACCAATAGCTATGTAACAGCGGAAAGCCCGTTGCAC 1578
Db      567  GlnAsnAlaValGlyThrAspLeuTyrGlnTrpAspAsnValProLysAlaValValGlu 586
QY      1579  GGTGTGGAATTCGCGCACATTGCG-----CTGTGTCAGAGGATGTC 1623
Db      587  GlyLeuGluGlySerLeuAsnValProValSerGluThrValMetTrpThrAsnAsnIle 606
QY      1624  ACCTGTCTCACTGAATTACACCTGGACCGCGAGTGAACACAGCTGATGGTGAACAAAGGT 1683
Db      607  ThrTyrMetLeu-----LysSerGluAsnLys-----ThrThrGly 618
QY      1684  GCGCCGCTGAGTTATACCCCTGAACATGTTGTAATGCGAAACTGAATGGCAGATCAC 1743
Db      619  AspArgLeuSerIleIleProGluTyrThrLeuAsnSerThrLeuSerTrpGlnAlaArg 638
QY      1744  GAAGAGTGCATCATGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1803
Db      639  GluAspLeuSerMetGlnThrPheThrTrpTyrGlyGlnGlnProLysLysTyr 658
QY      1804  AATTATTCGTCACTGAGCCCTGTACAGAAAGAGTGTATGATGAGAAAGGAGATACCTG 1863

```

```

Db      659  AsnTyrLysGlyGlnProAlaValGlyProGluThrLysGlu-----Ile 673
QY      1864  AAAGCCTGGACGGTGGTGGATGCGAGTCTGTCGTGGAAGATGACGGATGCCCTGACGCTG 1923
Db      674  SerProTyrSerIleValGlyLeuSerAlaThrTrpAspValThrLysAsnValSerLeu 693
QY      1924  AATGCTGGGTGAATAACCTGCTCAACAAGGATTACAGTGACGTGACGGCTGTACAGTGCC 1983
Db      694  ThrGlyGlyValAspAsnLeuPheAspLys-----ArgLeuTrpArgAla 708
QY      1984  GGTAAAGAGTACGCTG-----TATGCCGGTGTACTTCCAGACGGGATCATCAACA 2034
Db      709  GlyAsnAlaGlnThrThrGlyAspLeuAlaGlyAlaAsnTyrIleAlaGlyAlaGlyAla 728
QY      2035  ACAGGATATGTATACCTGACGCGAAATTTACTGGATGTGCTGAACTATCATGTTTC 2088
Db      729  TyrThrTyrAsnGluProGlyArgThrTrpTyrMetSerValAsnThrHisPhe 746

```

Search completed: October 13, 2004, 11:55:51
Job time : 86 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: October 13, 2004, 11:54:23 ; Search time 136 Seconds
(without alignments)
9995.325 Million cell updates/sec

Title: US-10-625-972-4
Perfect score: 3732
Sequence: 1 atggaataaacactctggc.....cgctgaactatcagttctga 2091

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1351062 seqs, 321799191 residues

Total number of hits satisfying chosen parameters: 2702124

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10625972/runat_12102004_130108_26931/app_query.fasta_1.2247
-DB=Published Applications AA -QMT=fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10625972.cgn_1.180.@runat_12102004_130108_26931
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:
1: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pcp.*
2: /cgn2_6/prodata/1/pubpaa/PCT_NEW_PUB.pcp.*
3: /cgn2_6/prodata/1/pubpaa/US06_NEW_PUB.pcp.*
4: /cgn2_6/prodata/1/pubpaa/US06_PUBCOMB.pcp.*
5: /cgn2_6/prodata/1/pubpaa/US07_NEW_PUB.pcp.*
6: /cgn2_6/prodata/1/pubpaa/PCTUS_PUBCOMB.pcp.*
7: /cgn2_6/prodata/1/pubpaa/US08_NEW_PUB.pcp.*
8: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pcp.*
9: /cgn2_6/prodata/1/pubpaa/US09A_PUBCOMB.pcp.*
10: /cgn2_6/prodata/1/pubpaa/US09B_PUBCOMB.pcp.*
11: /cgn2_6/prodata/1/pubpaa/US09C_PUBCOMB.pcp.*
12: /cgn2_6/prodata/1/pubpaa/US09_NEW_PUB.pcp.*
13: /cgn2_6/prodata/1/pubpaa/US10A_PUBCOMB.pcp.*
14: /cgn2_6/prodata/1/pubpaa/US10B_PUBCOMB.pcp.*
15: /cgn2_6/prodata/1/pubpaa/US10C_PUBCOMB.pcp.*
16: /cgn2_6/prodata/1/pubpaa/US10_NEW_PUB.pcp.*
17: /cgn2_6/prodata/1/pubpaa/US60_NEW_PUB.pcp.*
18: /cgn2_6/prodata/1/pubpaa/US60_PUBCOMB.pcp.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
------------	-------	-------------	-----------	----	-------------

Alignment Scores:	2.66e-78	Length:	682
Pred. No.:	972.00	Matches:	244

ALIGNMENTS

RESULT 1

US-10-238-075-1077
; Sequence 1077, Application US/10238075
; Publication No. US20030148324A1

GENERAL INFORMATION:

APPLICANT: I.N.S.E.R.M.

TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolat

FILE REFERENCE: BLANDINE

CURRENT APPLICATION NUMBER: US/10/238.075

PRIOR FILING DATE: 2002-09-10

PRIOR APPLICATION NUMBER: 0003145

PRIOR FILING DATE: 2000-03-10

NUMBER OF SEQ ID NOS: 1576

SOFTWARE: Patent in version 3.1

SEQ ID NO 1077

LENGTH: 682

TYPE: PRT

ORGANISM: Escherichia coli

US-10-238-075-1077

1	972	26.0	682	14	US-10-238-075-1077	Sequence 1077, Ap
2	428.5	11.5	687	12	US-10-282-122A-48522	Sequence 48522, A
3	399	10.7	684	12	US-10-282-122A-48915	Sequence 48915, A
4	388	10.4	660	14	US-10-238-075-732	Sequence 732, App
5	371	9.9	698	12	US-10-282-122A-77625	Sequence 77625, A
6	355.5	9.5	654	14	US-10-238-075-363	Sequence 263, App
7	347.5	9.3	713	12	US-10-238-075-527	Sequence 527, App
8	321	8.6	713	12	US-10-282-122A-77578	Sequence 77578, A
9	302.5	8.1	774	12	US-10-282-122A-48409	Sequence 48409, A
10	302	8.1	729	12	US-10-282-122A-56053	Sequence 56053, A
11	301	8.1	774	12	US-10-282-122A-75216	Sequence 75216, A
12	302	8.1	763	12	US-10-282-122A-48639	Sequence 48639, A
13	300.5	8.1	774	12	US-10-282-122A-43232	Sequence 43232, A
14	295	7.9	722	12	US-10-282-122A-55554	Sequence 55554, A
15	293.5	7.9	755	12	US-10-282-122A-69342	Sequence 69342, A
16	293	7.9	718	12	US-10-282-122A-44987	Sequence 44987, A
17	290.5	7.8	696	12	US-10-282-122A-76306	Sequence 76306, A
18	283.5	7.6	769	12	US-10-282-122A-48576	Sequence 48576, A
19	283.5	7.6	784	12	US-10-282-122A-66570	Sequence 66570, A
20	283.5	7.6	825	12	US-10-282-122A-51169	Sequence 51169, A
21	281	7.5	699	12	US-10-282-122A-77914	Sequence 77914, A
22	278	7.4	747	12	US-10-282-122A-43354	Sequence 43354, A
23	272	7.3	820	12	US-10-282-122A-66383	Sequence 66383, A
24	270.5	7.2	735	12	US-10-282-122A-60017	Sequence 60017, A
25	269	7.2	687	14	US-10-181-319-2	Sequence 2, Appli
26	267.5	7.2	813	9	US-09-815-242-5085	Sequence 5085, Ap
27	267.5	7.2	813	12	US-10-282-122A-43497	Sequence 43497, A
28	267.5	7.1	704	12	US-10-282-122A-67879	Sequence 67879, A
29	264	7.1	743	12	US-10-282-122A-69539	Sequence 69539, A
30	263.5	7.1	252	11	US-09-764-868-710	Sequence 710, App
31	263	7.0	274	9	US-09-764-868-707	Sequence 707, App
32	263	7.0	774	12	US-10-282-122A-68080	Sequence 68080, A
33	250.5	6.7	926	15	US-10-418-861B-18	Sequence 18, Appl
34	247.5	6.6	767	12	US-10-335-977-4925	Sequence 4925, Ap
35	247.5	6.6	812	12	US-10-282-122A-68347	Sequence 68347, A
36	246	6.6	669	12	US-09-809-665A-105	Sequence 105, App
37	243	6.5	721	12	US-10-282-122A-69343	Sequence 69343, A
38	242.5	6.5	727	12	US-10-282-122A-49051	Sequence 49051, A
39	241.5	6.5	727	12	US-10-282-122A-49650	Sequence 49650, A
40	240.5	6.4	228	9	US-09-764-868-1123	Sequence 1123, Ap
41	240.5	6.4	228	11	US-09-764-875-1022	Sequence 1022, Ap
42	239	6.4	793	12	US-10-335-977-5080	Sequence 5080, Ap
43	239	6.4	793	12	US-10-335-977-5081	Sequence 5081, Ap
44	236	6.4	970	14	US-10-331-061-5	Sequence 5, Appli
45	236	6.3	791	12	US-10-282-122A-69538	Sequence 69538, A

Percent Similarity:	51.08%	Conservative:	135
Best Local Similarity:	32.88%	Mismatches:	245
Query Match:	26.05%	Indels:	118
DB:	14	Gaps:	23
US-10-625-972-4 (1-2091) x US-10-238-075-1077 (1-682)			
QY	7	ATAACCACTCGTTCGCTCGAGTATTCCTCCCTGCTCTCGATTTTCAGCCAGCATAGCT	66
DB	11	IleAlaValMetCysSerAlaValIleSerSer---GlyTyrAlaSerSerAsp-----	27
QY	67	GCTCCAGAGATGATGATGCTCGCATCCGGTATGAGAAAGCTGATACAGCA	126
DB	28	LysIysGluAspThrLeuValValThrAlaSerGlyPheThrGlnGlnLeuArgAsnAla	47
QY	127	GCCGCCAGTGTTCGTGATTACAGGAGGAATTCAGTCCAGCCAGTACACAGATCTG	186
DB	48	ProIaSerValSerValIleThrSerGluGlnLeuGlnLysLysProValSerAspLeu	67
QY	187	GCGAGGCTCTGATCAGTACAGTACAGGCTGATGTTGAAAGTGGTACGGGTAAACCCGA	246
DB	68	ValAspAlaValLysAspValGluGlyIleSerIleThrGlyAsnGluLysPro---86	
QY	247	GGCTGGAATCAGCATCCAGAGATGCGAGTACAGCCAGTACAGCTGATGATGCT	306
DB	87	-----AspIleSerIleArgGlyLeuSerGlyAspTyrThrLeuIleLeuValAspGly	104
QY	307	GTTCTGCTCAGCGGCGAGCAGTACGCTCCACAGCGTTCCTGCCATGAATACCGGG	366
DB	105	ArgArgGlnSerGly---ArgGluSerArgProAsnGlySerGlyGlyPheGluAlaGly	123
QY	367	TTATGCCCTCTGCGCCGATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGT	426
DB	124	PheIleProValGluAlaIleGluArgIleGluValIleArgGlyProMetSerSer	143
QY	427	CTGTATGGCTCATGCGATGGCGGTGTGCTGAATCATATTACACAGAAAG---AATGCA	483
DB	144	LeuTyrGlySerAspAlaIleGlyValIleAsnIleIleThrLysProValAsnAsn	163
QY	484	GACAAATGGCTCTTCCTCGCTCAATGAGGCTGAATCTCAGAGAAAGCAAAATGGGT	543
DB	164	GlnThrTyrAspGlyValLeuGlyLeuGlyValIleIleGlnGluHisGlyLysPheGly	183
QY	544	AACAGCAGCCAGTTAATTTCTGGACAGTGTCCCTTTGTGATGATGATGATGATGATG	603
DB	184	AsnSerThrAsnAspPheTyrLeuSerGlyProLeuIleLysAspLysLeuGlyLeu	203
QY	604	CAGGTACGGGTAGCACACACAGCGTCAAGGTTTCATCGCTCACATCACTGAGCGATACA	663
DB	204	GlnLeuTyrGlyGlyMetAsnTyrArgLysGluAspSerIleSerGln-----219	
QY	664	GCAGGCACGGTATCTCTTATCCCGGAGTACAGAAATTAATCTGGTGCAGCTCTT	723
DB	220	---GlyThr-----ProAlaLysAspAsnLysIleThrAlaThrLeu	233
QY	724	GACTGGAAGCGCGGAGCAGGATGCTCTGTTGATGATGATGATGATGATGATGATGATG	783
DB	234	GlnPheThrProThrGluSerGlnLysPheValPheGluTyr-----247	
QY	784	TATGATAACGGGATGGCACTGGGGAGTCTGAGGGGGGATGATGACCGGACCTGGCG	843
DB	248	-----GlyLysAsnAsnGlnValHisThrLeuThrProGlyGluSerLeuAspAlaThr	265
QY	844	TATGACGGAACAAATTTTCAGTGGCTGATGATCATCTTTCACCTTCGGACATGGAAA	903
DB	266	ThrMetArgGlyAsnLeuLysGlnProAsnSerLysArgGluThrHisAsnSerArgSer	285
QY	904	TGATATCTG---AATGGGAAC-----GAGACAGAAATAAA	936
DB	286	HisTrpValAlaAlaTrpAsnAlaGlnGlyGluIleLeuHisProGluIleAlaValTyr	305
QY	937	GGTCTGTGACCTGTACGAGTGTACTG-----AAGCGGCAAA-----TGGGGG	981

DB	306	GlnGluLysValIleArgGluValLysSerGlyLysLysAspLysTyrAsnHisTrpAsp	325
QY	982	CTTGCCGCTCAGCCGCGGAGCTTAAGAAATCGAACCTTATCCTGAATTAATTAATCTG	1041
DB	326	LeuAsnTyrGluSerArgLysProGluIleThrAsnThrIleIleAspAlaLysValThr	345
QY	1042	ACCCCTCTGGGAGAAATCTCATCTGTTACGGTGGGGCGGAGTTTCAGAGTCTGCTCATG	1101
DB	346	AlaPheLeuProGlu---AsnValLeuThrIleGlyGlnPheGlnHisAlaGluLeu	364
QY	1102	AAAGACGGAGTTGCTCTTCGCCAGCACAGGTGAA-----ACT	1137
DB	365	ArgAspAsp-----SerAlaThrGlyLysLysThrThrGluThrGlnSerValSer	381
QY	1138	TTCGGCAGAAAGCTGCTCGTATTTCTGAGGATGAGTGGCATCTCAGGATGCACTT	1197
DB	382	IleLysGlnLys-----AlaValPheIleGlnGluThrAlaAlaThrAspSerLeu	399
QY	1198	GCGTGAATCGGGCAGCCGCTATGAACATCATGAGCAATTCGGGGGACACTTCAGTCCG	1257
DB	400	AlaLeuThrGlyGlyLeuArgLeuAspAsnHisGluIleTyrGlySerTyrTrpAsnPro	419
QY	1258	CGTGCATATCTGGTCTGGGATGTCGAGATGCCCTGAGCGCTGAAGCGGTGACACAG	1317
DB	420	ArgLeuTyrAlaValTyrAsnLeuThrAspAsnLeuThrLeuLysGlyGlyIleAlaLys	439
QY	1318	GCATATAAGCACCAGAAATGGGCGCATACATAAAGGATTAGTGTGTCTCGGGCAG	1377
DB	440	AlaPheArgAlaProSerIleArgGluValSerProGlyPheGlyThrLeuThrGlnGly	459
QY	1378	GGAAAAACAATCTACTTGTGTAAACCCGCTGAGCGGAGCGGAGAGCGTCATGATGAG	1437
DB	460	GlyAlaSerIleMetTyrGlyAsnArgAspLysProGluThrSerValThrGluGlu	479
QY	1438	GCTGGGTGATTACGATAAAGCGCGCTGCAATGCCAATGTCACAGGTTTATGACT	1497
DB	480	IleGlyIleIleTyrSerAsnAspSerGlyPheSerAlaSerAlaThrLeuPheAsnThr	499
QY	1498	GACTTCTCCAAAGATGCTCTTATTCATTAATGAT-----1536	
DB	500	AspPheLysAsnLysLeuThrSerTyrAspIleGlyThrLysAspProValThrGlyLeu	519
QY	1537	AACACCAATAGCTATGTAACACGCGGAGGCGCGGTGTCACGCTGTGGAATTTGCGGC	1596
DB	520	AsnThrPheIleTyrAspAsnValGlyGluAlaAsnIleArgGlyValGluLeuAlaThr	539
QY	1597	ACATTGCGCTGTGGTCAGAGGATGTACGCTGTCACTGAATTAACCTGAGCCCGAAGT	1656
DB	540	GlnIleProValTyr---AspLysTyrHisValSerAlaAsnTyrThrPheThrAspSer	558
QY	1657	GAACAACGTGATGGTGATAAC-----AAAGTGGCGCGCTGAGT	1695
DB	559	ArgArgLysSerAspAspGluSerLeuAsnGlyLysSerLeuLysGlyGluProLeuGlu	578
QY	1696	TATACCCCTCAACACATGTTGAATGCGAACTCAACTCACTGAGATCACCGAAGAGTGCGA	1755
DB	579	ArgThrProArgHisAlaAlaAsnAlaLysLeuGluTyrAspTyrThrGlnAspIleThr	598
QY	1756	TCATGGCTGGTGGCGCTTATCCGGG-----1782	
DB	599	PheTyrSerSerLeuAsnTyrThrGlyLysGlnIleTrpAlaAlaGlnArgAsnGlyAla	618
QY	1783	AAAAACACAGCTTTCACCCAGAAATTTCTGCTACTGAGCGCTGTACAGAAAGAGTGTAT	1842
DB	619	LysValProArgValArgAsnGlyPheThrSerMet-----630	
QY	1843	GATGAGAAAGAGATACCTGTAAGCGCTGAGCGGTGGTGGTGGTGGTGGTGGTGGTGGT	1902
DB	631	-----AspIleGlyLeuAsnTyrGln	637
QY	1903	ATCAGCGATGCCCTGACGCTGAATGCTGGGTGAATAACCTGCTCAACAGGATTACAGT	1962
DB	638	IleLeuProAspThrLeuIleAsnPheAlaValLeuAsnValThrAspArgLysSerGlu	657

QY 1963 GACGTGACGCTGTACAGTCCGGTAAGAGTACGCTGTGTATGCGGTGATTACTTCCAGAGC 2022
 Db 658 AspileAspThrileAsp-
 QY 2023 GGATCATCAACACAGGATATGTATACCTGAGCGAAATTTACTGTGATGTCCTGCAACTAT 2082
 Db 667 -----GlnValAspGluGlyArgAGTyrTTPAlaAsnValArgVal 680

QY 2083 CAGTTC 2088

Db 681 SerPhe 682

RESULT 2

US-10-282-122A-48522
 ; Sequence 48522, Application US/10282122A
 ; Publication NO. US20040029129A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Liangsu
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Malone, Cheryl
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari
 ; APPLICANT: Zyskind, Judith
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John
 ; APPLICANT: Cart, Grant
 ; APPLICANT: Yamamoto, Robert
 ; APPLICANT: Forsyth, R.
 ; APPLICANT: Xu, H.
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A
 ; CURRENT APPLICATION NUMBER: US/10/282,122A
 ; CURRENT FILING DATE: 2003-02-20
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/230,335
 ; PRIOR FILING DATE: 2000-09-06
 ; PRIOR APPLICATION NUMBER: 60/230,347
 ; PRIOR FILING DATE: 2000-09-09
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/267,636
 ; PRIOR FILING DATE: 2001-02-09
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 78614
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 48522
 ; LENGTH: 687
 ; TYPE: PRT
 ; ORGANISM: Bacteroides fragilis

US-10-282-122A-48522
 Alignment Scores:
 Pred. No.: 2,81e-29
 Score: 428.50
 Percent Similarity: 40.00%
 Best Local Similarity: 24.18%
 Query Match: 11.48%
 DB: 12

Length: 687
 Matches: 191
 Conservative: 125
 Mismatches: 261
 Indels: 213
 Gaps: 38

US-10-625-972-4 (1-2091) x US-10-282-122A-48522 (1-687)

QY 16 CTGGCTTCGCTAGTATCCCTCTCTCGGATTTTCCAGCCAGCAGCATAGCTGCTGCAGAG 75
 Db 8 LeuLeuLeuThrLeuLeuGlyThrLeuProLeuAlaAlaGlnAsnValArgGlnGluGln 27
 QY 76 GATGTG-----ATGATTCTCTCGGCA 96
 Db 28 AspThrValSerTyrMetAsnAspAspProPheAsnLeuGluGlnLeuValValThrAla 47
 QY 97 TCCGGCTATCAGAAAAAGCTGACTAAACGACGCGCCAGTGTCTCTCTGTTTATGATCCAGGAG 156
 Db 48 ThrArgThrGluLysLysLysLysLysAsnThrProValThrGlnLeuThrSerLys 67
 QY 157 GAATTGCAGTCCAGCCAGTACACAGTCTCTGGCGGAGCTCTG---AGATCAGTAGAGGT 213
 Db 68 GlnIleGluArgGlyThrGlyAsnIleGlnAspLeuLeuThrGlnGluValProGly 87
 QY 214 GTGGATGTTGAAAGTGTGCTACGGGTAAACCGGA---GGGCTGGAAATACGATCCGAGGA 270
 Db 88 LeuAsnPheGln-----GluValGlyTyrGlyThrSerLeuAspIleGlnGly 103
 QY 271 ATGCCAGCCAGTTACACGCTGATATCTGATGATGTTGTTCTG-----CAGGGC 318
 Db 104 LeuGlySerLysHisIleLeuPheLeuIleAspGlyGluArgIleAlaGlyGluAsnGly 123
 QY 319 GGAAGCAGTACGCTGACTCCCAACGGTTTTCTGCCATGAATACCGGGTTCATGCCCT 378
 Db 124 GlyAsnIleAsp-----TyrSerArgIleAsn-----132
 QY 379 CTGGCCGCCATTGAGCTATTGAGTTATCAGGGGCGCGATGTCACACTGTATGCTCT 438
 Db 133 LeuTyrAsnIleAspHisIleGluIleValLysGlyAlaSerSerAlaLeuTyrGlySer 152
 QY 439 GATCGATGGCGGTGTGTGATATCAATACAGAAAGATGCAGACAAATGGCTCTCT 498
 Db 153 GlnAlaMetGlyGlyValIleAsnIleIleThrArgLysAlaLysLysPhe-----170
 QY 499 TCCGTCAATGACGGCTGAATCTGCAGAAAGCAAAATGGGGTAAACAGCAGCAGTTT 558
 Db 170 -----170
 QY 559 AATTTCTGGAGCAGTGTGCTCCCTTGTGGATGATCTCTCAGCTGAGTACGGGTAGC 618
 Db 171 -----GluAlaSerAlaGlyIleArgTyrAlaGlyArg 181
 QY 619 ACACAACAGCGTCAGGGTTCATCGGTACATCATCAGCATACAGCAGCAGCGTATT 678
 Db 182 AsnGlnGln-----AsnTyrLysAspThrProLysAspHisSer 194
 QY 679 CCTTAT-----CCCACGGAGTCACAGAAATTAATCTTGGT 714
 Db 195 GlnTyrLysTyrArgIleHisLeuAspLysProAsnLeuAsnThrAsnLeuSerLeuGly 214
 QY 715 GCAGTCTTCTGCTGGAAGGCGTCCGACGAGATGCTCTGTTTGTATGATGATACACC 774
 Db 215 LeuAsnLeuGlyLysPheThrMetAsnThrAspValLeuTyrLysSerPheAsp---Gly 233
 QY 775 CGCAGCGTATTATGATAACCGG-----795
 Db 234 TyrGlnLeuPheAspLysLysProLeuValLysTyrPheProAlaTyrAsnThrThrIle 253
 QY 796 -----GATGGCACTGGG 810
 Db 254 ThrGluGluLeuSerLysProProThrSerIleSerGlyTyrGluAspValGlnValAla 273
 QY 811 -----AGTCTGACGGGGGATATGAC 831
 Db 274 HisLysMetAspTyrArgPheSerLysArgLeuLysValGlnLeuLysGlySerTyr---292
 QY 832 CGGACCCCTGCTGCTATGAGCGAAACAAATTTTCACTGGCTATGATCATCTTTC-----885
 Db 293 -----TyrMetLeuAsnLysTyrAspPheGlnAlaAspAsnIlePheGluLys 308
 QY 886 -----ACCTTCGGAAACATGGAAATCGTATCTGAACCTGGAACGAGCAGAGAAAT 933

```

Db      309 SerGluAspTyrThrThrGlyGlySerIleAspTyr-----ThrIleSer 323
Qy      934 AAAGGTCGTGAGCTGTGACGACAGTGTA-----CTGAAGCGCGCAAAATGG 978
Db      324 AspLysSerSerLeuValAlaSerValHisThrAspHisTyrAsnArgTyrAspLysTyr 343
Qy      979 GGGCTT---CCCGGTGACCGCGGAGCTTAAGAAATCGAACCTTATCTCGTAATTCATTA 1035
Db      344 GluLeuLysSerGlyArgLeuGluTyrLysAsnAsnIleIleGlnProArgIleVal 363
Qy      1036 CTGCTTACCCCTCTGGGAGAAATCTCATCTGTTACCGTGGCGGCGAGTTTCAGAGCTCG 1095
Db      364 TyrSerThrThrAlaLeuAspLysGlnThrIleThrGlyGlyLeuGluTyrTyrArgGlu 383
Qy      1096 TCCATGAAGACGGAGTGTCTTCCGACGACAGGTGAACCTTCCGCGCAGCAAAAGCTGG 1155
Db      384 SerLeuPheSerAspLysPhe-----GluThrGlyValLysGluAsnLysSerGlnTrp 401
Qy      1156 -----TCGGTATTTCGAGGATGAGTGCATCTCACGGATGCACCTGGCTGACTGG 1209
Db      402 TyrAlaThrAlaPheLeuGlnAspAspTrpSerIleAsnLysGlnPheSerValIleAla 421
Qy      1210 GGCAGCGCTATGAACATCATGAGCAATTCGGGGACACTTCAGTCCGCGTCATATCTG 1269
Db      422 GlyLeuArgCysAspTyrHisGluLysTyrGlyThrAsnLeuThrProLysAlaSerVal 441
Qy      1270 GTCTGGGATGGCAGATGCTGGAGCTCAAGGCGGTGTGACACCGGGATATAAGCA 1329
Db      442 MetTyrLysIlePhe---ProPheThrValArgPheAsnTyrAlaArgGlyTyrArgSer 460
Qy      1330 CCAGAAATGGCGGATACATAAGGATTAGTGGTGTCCGGCGAGCAAAACAAT 1389
Db      461 ProSerIleLysGluLeuTyrMetAsnTrpAspHisLeu-----GlyMetPheTrp 477
Qy      1390 CTACTTGGTAACCCGACCTGAAGCGGAGAGAGC-----GTCAGTATGAGCTGGG 1443
Db      478 IleTyrGlyAsnSerLysLeuLysProGluThrAsnAsnTyrIleSerLeu---SerGly 496
Qy      1444 GTGTATTACCATACCCCGCGCTCGAATGCCAATGTCACAGGTTTATGACTGACTTC 1503
Db      497 GluTyrValAsnSerTrpIleAsnIleAsnAlaAsnVal-----TyrSerAsnTrpPhe 514
Qy      1504 TCCAAAGATGTTCTCTTATTCCTAATGATATAACACCAAT---AGCTATGTAACAGC 1560
Db      515 ArgAsnLysIleGluGlyMetTrpSerAsnAspGlnThrGluLeuHisTyrIleAsnIle 534
Qy      1561 GGAAGCGCGGTTGACCGTGTGGAATTTCCGGGCACATTCGCGTGGTGCAGAG--- 1617
Db      535 GlyLysSerArgLeuAlaGlyValGluThrMetCysLysIleGlnIleAsnArgHisIle 554
Qy      1618 GATGTCACGCTGTCATGATTAACCTGGACCCGGAAGTGAACAAAGTGGTGGTATGATAAC 1677
Db      555 AsnValHisGlyAlaTyrAsnTyrLeuTyrThr-----SerLysAspAlaAsp--- 570
Qy      1678 AAAGTGGCCGCTGATGTTATACCCCTGACACATGGTGAATGCGAACTGAACCTGGCAG 1737
Db      571 ---GlyValArgLeuSerSerSerProHisSerGlyAsnIleArg----- 585
Qy      1738 ATCACCAGAGAGGTGCATCATGCTGGGTGGTGGTTCGCGGGAACACACCACTGTTTC 1797
Db      586 -----AlaGluTyrAsnThrArgIleProArgTyr 595
Qy      1798 ACCAG-----AATTATTGTCATGACGCGCTGTACAGAGAAA-----GTGTATGAT 1845
Db      596 AlaThrValValAsnLeuSerGlyAsnIleMetGlyLysLysLysPheAspValLeuAsp 615
Qy      1846 GAG-----AAAGGAGAAATACCTGAAGCTCG-----ACGGTGGTGGATCAGCT 1890
Db      616 GluLeuGluIleAspLysLysValGluAlaTyrTyrGlnAlaLysValAsnProTyr 635
Qy      1891 CTGTCGTGGAAGATGAC-----GATGCCCTGACCGCTGAATCTCGG 1932

```

```

Db      636 CysLeuTrpAspLeuThrValSerGlnTyrIleMetGlnAsnLeuArgIleThrAlaGly 655
Qy      1933 CTGAATAACCTGCTCAACAAGGATTACAGTGACGTGACCGCTGTACAGTCCGGTAAAGAT 1992
Db      656 IleThrAsnLeuPheAsp-----TyrThrSerAspArgVal 667
Qy      1993 ACGCTGTATGCGGTGATTACTTCCAGACGGGATCATCAACACAGGATATGTGATACCT 2052
Db      668 Thr-----PheAsnThrSerThrSer-----Pro 675
Qy      2053 GAGCGAAATTACTGGATGTCGCTGAACAT 2082
Db      676 GlyArgAsnTyrPheIleAlaCysAsnTyr 685

```

RESULT 3

```

US-10-282-122A-48915
; Sequence 48915, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 48915
; LENGTH: 684
; TYPE: PRT
; ORGANISM: Bacteroides fragilis
US-10-282-122A-48915

```

```

Alignment Scores:
Pred. No.: 1,28e-26 Length: 684
Score: 399.00 Matches: 164
Percent Similarity: 42.57% Conservative: 131
Best Local Similarity: 23.67% Mismatches: 264
Query Match: 10.69% Indels: 134
DB: 12 Gaps: 34

```

```

US-10-625-972-4 (1-2091) x US-10-282-122A-48915 (1-684)

```


APPLICANT: I. N. S. E. R. M.
TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolable from the cells of the microorganism *Escherichia coli*.
TITLE OF INVENTION: *E. coli*, and biological uses of these polynucleotides and of their derivatives.

```

; FILE REFERENCE: BLANDINE
; CURRENT APPLICATION NUMBER: US/10/238,075
; CURRENT FILING DATE: 2002-09-10
; PRIOR APPLICATION NUMBER: 0003145
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 1576
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 732
; LENGTH: 660
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-238-075-732

```


QY 1624 -----ACGCTGTCACTGAATTACACCTGGACCCGCAAGTGAACAACGTGAT 1668
 Db 535 LysTyrThrThrAspLeuPheSerLeuAspValAlaTyrAsnArgThrArgGlyLysAsp 554
 QY 1669 GGTGATAACAAGGTGCGCCGCTGAGTTATACCCCTGAACACATGGTGAATGCGAAACTG 1728
 Db 555 ThrAspThr--GlyGlutYrIleSerSerIleAsnProAspThrValThrSerThrLeu 573
 QY 1729 AACTGGCAGATCACCGAAGAGTGGCATCA-----TGGCTGGGTGCGCGTTATCGCGGG 1782
 Db 574 AsnIleProIleAlaHisSerGlyPheSerValGlyTrpValGlyThr---PheAlaAsp 592
 QY 1783 AAAACACACCGGTTCCACCCAGAATTATTCGTCACTGACGCGTGTACAGAAGAAAGTGTAT 1842
 Db 593 ArgSerThrHisIleSerSerSerTyrSerLys----- 603
 QY 1843 GATGAGAAGAGAGATACTGAAGCCCTGGACGGTGGTGGATGCGAGGTCTCTGCTGGAG 1902
 Db 604 -----GlnProGlyTyrGlyValAsnAspPheTyrValSerTyrGln 617
 QY 1903 ATGACGGATGCGCTG-----ACGCTGAATGCTGGCGGTGAATAACCTGCTCAACAAG 1953
 Db 618 GlyGlnGlnAlaLeuLysGlyMetThrThrThrLeuValLeuGlyAsnAlaPheAspLys 637
 QY 1954 GATTAAC 1959
 Db 638 GluTyr 639

RESULT 5

US-10-282-122A-77625
 ; Sequence 77625, Application US/10282122A
 ; Publication No. US20040029129A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Liangsu
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Malone, Cheryl
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Chilsen, Kari
 ; APPLICANT: Zyskind, Judith
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John
 ; APPLICANT: Carr, Grant
 ; APPLICANT: Yamamoto, Robert
 ; APPLICANT: Forsyth, R.
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
 ; FILE REFERENCE: ELITRA.034A
 ; CURRENT APPLICATION NUMBER: US/10/282,122A
 ; PRIORITY FILING DATE: 2003-02-20
 ; PRIORITY FILING DATE: 2000-03-21
 ; PRIORITY FILING DATE: 2000-03-21
 ; PRIORITY FILING DATE: 2000-05-23
 ; PRIORITY FILING DATE: 2000-05-23
 ; PRIORITY FILING DATE: 2000-05-26
 ; PRIORITY FILING DATE: 2000-09-06
 ; PRIORITY FILING DATE: 2000-09-06
 ; PRIORITY FILING DATE: 2000-09-09
 ; PRIORITY FILING DATE: 2000-12-22
 ; PRIORITY FILING DATE: 2000-10-23
 ; PRIORITY FILING DATE: 2000-11-27
 ; PRIORITY FILING DATE: 2000-11-27
 ; PRIORITY FILING DATE: 2000-12-22
 ; PRIORITY FILING DATE: 2001-02-09
 ; PRIORITY FILING DATE: 2001-02-09
 ; PRIORITY FILING DATE: 2001-02-16
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 78614
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 77625

; LENGTH: 698
 ; TYPE: PRT
 ; ORGANISM: Vibrio cholerae
 US-10-282-122A-77625
 Alignment Scores:
 Pred. No.: 4,34e-24 Length: 698
 Score: 371.00 Matches: 176
 Percent Similarity: 41.59% Conservative: 143
 Best Local Similarity: 22.95% Mismatches: 292
 Query Match: 9.94% Indels: 156
 DB: 12 Gaps: 35

US-10-625-972-4 (1-2091) x US-10-282-122A-77625 (1-698)

QY 16 CTGCTTCGCTAGTATCCCTGCTCGGATTTTCAGCC---AGCAGCATAGCTGCTGCTCA 72
 Db 12 LeuSerSerAlaIleMetLeuAlaValProSerAlaTyrAlaAspThrAlaSer 31
 QY 73 GAGGATGTGATGATTCCTCGCATCCGCTATGAGAAAAGCTGACTAACGCGCGGCC 132
 Db 32 PheAspGluValValSerThrArgLeuAsnThrGlnIleThrAspThrAlaAla 51
 QY 133 AGTGTTCCTGATAGCCAGGAGGAAATGCCAGTCCAGCCAGTACCACCATCTGGCGGAG 192
 Db 52 SerValAlaValIleAsnAlaSerAspIleGluGlnGlnMetAlaGluAspIleGluGly 71
 QY 193 GCTCTGAGATCAGTAGAGGGTGTGGATGTTGAAAGTGGTACCGGTAAACCCGAGGGCTG 252
 Db 72 LeuPheLysTyrThrProGlyValThrLeuThrThrAsnSer---ArgGlnGlyValGln 90
 QY 253 GAAATCAGCATCCGAGGATGCCAGCCAGTACACGCTGATCTACTGATTGCTGCTCGT 312
 Db 91 GlyIleAsnIleArgGlyIleGluGlyAsnArgIleLysValIleValAspGlyValAla 110
 QY 313 CAGGCGGGAAGCAGTACGCTGCTCCCAACGGTCTTCTGCCATGAATACCGGTTTCATG 372
 Db 111 GlnProAsnGlnPheAsp-----SerGlyAsnSerPheLeuAsnSerSerArgVal 127
 QY 373 CCC---CCTCTGGCCGCATGAGCGTATTAGGTTATCAGGGGCGCGGATGTCCACTG 429
 Db 128 AspIleAspThrAspMetValLysSerValGluIleValLysGlyAlaAlaSerSerLeu 147
 QY 430 TATGGCTCTGATCGATGGCGGTGGTGAATATCATTCACAGAAAGCAAAATGCGACAAA 489
 Db 148 GlnGlySerAspAlaIleGlyIleValAlaPheGluThrLysAspProAlaAsp--- 166
 QY 490 TGGCTCTCTCCGCTCAATCGAGGCTGAATCTGCAGGAAAGCAAAATGGGTAAACAGC 549
 Db 167 -----IleLeuLysGlyArgAsnMet-----GlyGlyTyr 176
 QY 550 AGCAGTTTATTTCTGGACAGTGTCCCTTGTGATGATTCCTCAGCTGCAGGTA 609
 Db 177 AlaLysLeuAsnTyrSerSerAspLysThrPheSerIleAlaLeuAlaAsn 196
 QY 610 CGCGGT-----AGCACAAACAGCGTCAGGTTTCATCGGTCA 648
 Db 197 LysSerGlyAspLeuGluSerLeuValAlaTyrThrArgAspGlyGlnGluIleGln 216
 QY 649 TCACGTAGCGATACAGCAGCAGCGGTATTCCTTATCCACGAGGATCAGAGATTATAT 708
 Db 217 AsnPheGlySer-----ProAspGlnGlnAspAsnAlaAsnAsn 230
 QY 709 CTTGGTGCAGTCTTGACTGGAAGCGTGGAGCAGGATGTGCTGTTTGTATATGAT 768
 Db 231 LeuLeuValLysLeuGlnTyrGlnLeuAsnProLysHisArgLeuGluPheSerGlyAsn 250
 QY 769 ACCACCGCGCAGGTTATGATAACCGGATGGCAACTGGGGAGTCTG-----ACGGGG 822
 Db 251 TyrIleArgAsnLysAsnAspLeuGluAsnLeuGluPheSerGlyTyrLysAsnAlaSer 270
 QY 823 GGATATGACCGGACCTGCGCTATGAGCGAAACAAATTTTCAGTGGCTGATCATACT 882

543 -----AsnArgThrGluSerMetAspAlaSerSerArgGluThrLeuGlySerGlyValAsn 560
 1702 CCTGAACACATGGTGAATGCGAAACTGAATGCGAGATCACCGAAGAGGTGGCATATGG 1761
 561 ProAspThrLeuIleSer-----AspIleSerIlePro 571
 1762 CTGGGTGCGCGT-----TATCGCGGGAACACACACCGTTTACCCAGATTATTCGTCA 1815
 572 ValGlyHisArgGlyValTyrAlaGlyTyrParg-----AlaGlu 584
 1816 CTGACGCGCTACAGAGAAAGTGTATGATGAGAAAGGAGAA---TACCTGAAAGCCTGG 1872
 585 LeuSerAlaSerAlaThrHisVal-----LysLysGlyAspProHisGlnAlaGlyTyr 602
 1873 ACGGTGGTGATGAGGTCTCTGTCGAGATGAGCGATGCGCTGCAATGCTGCG 1932
 603 ThrIleHisSerPheSerLeuSerTyrLysProValSerValLysGlyPheGluAlaSer 622
 1933 GTG-----AATAACCTCTCAACAGAGATTACAGTGACGTGACGTGCGCGGT 1986
 623 ValThrLeuAspAsnAlaPheAsnLys-----LeuAlaMetAsnGly 636
 1987 AAGAGTACGTGTATGCGGTGATTACTTCCAGACGGGATCATCAACACAGGATAT 2043
 637 LysGlyValProLeuSerGly-----ArgThrValSerLeuTyrThrArgTyr 652

RESULT 8

US-10-282-122A-77578
 ; Sequence 77578, Application US/10282122A
 ; Publication No. US20040029129A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Liangsu
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Malone, Cheryl
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari
 ; APPLICANT: Zyskind, Judith
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John
 ; APPLICANT: Yamamoto, Robert
 ; APPLICANT: Forsyth, R.
 ; APPLICANT: Xu, H.
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
 ; FILE REFERENCE: ELITRA.034A
 ; CURRENT APPLICATION NUMBER: US/10/282,122A
 ; CURRENT FILING DATE: 2003-02-20
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/230,335
 ; PRIOR FILING DATE: 2000-09-06
 ; PRIOR APPLICATION NUMBER: 60/230,347
 ; PRIOR FILING DATE: 2000-09-09
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/267,636
 ; PRIOR FILING DATE: 2001-02-09
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 78614
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 77578
 ; LENGTH: 713
 ; TYPE: PR1

ORGANISM: Vibrio cholerae
 US-10-282-122A-77578
 Alignment Scores:
 Pred. No.: 5,576-20 Length: 713
 Score: 325.50 Matches: 193
 Percent Similarity: 38.43% Conservative: 116
 Best Local Similarity: 24.00% Mismatches: 264
 Query Match: 8.72% Indels: 231
 DB: 12 Gaps: 48
 US-10-625-972-4 (1-2091) x US-10-282-122A-77578 (1-713)
 QY 1 ATGCGAATAACCACTCTGCTTCCGTAGTCAATTCCTGTCGGATTTCAGCCAGCAGC 60
 DB 1 MetLysLeuSerProValSerAlaAlaValLeuSerValLeu-----AlaAlaGly 17
 QY 61 ATAGCTGCTGCACAG-----GATGTGATGATGCTTCGCGATCCGCGC 102
 DB 18 PheAlaHisAlaGluThrGluProSerHisTyrGluGluValValValAlaAsnArg 37
 QY 103 TATGAGAAAAGCTGACTAAACGCGCAGCGCGAGTGTCTGTGATTAGCCAGAGGAATG 162
 DB 38 IleGluGlnProLeuSerGluValAlaGlySerValAlaValLeuGluGlyGluThrLeu 57
 QY 163 CAGTCCAGCCAGTACCAGCATCTGCGGAGGCTCTGAGATCAGTAGAGGTGTGATGTT 222
 DB 58 GluLysGlnGlyLysThrGluLeuTyrAspAlaLeuAsnGlnGluProGlyValSerVal 77
 QY 223 GAAAGTGTACGGGTAAACCGGAGGCTGGAATCAGCATCCGAGGAATGCCAGCAGT 282
 DB 78 ThrGlyGlyAlaGlyArgPro-----GlnAsnIleThrIleArgGlyMetThrGlyAsn 95
 QY 283 TACACGCTGATCTACTGATTGATGTTCTGTCAGGGC-----GGAAGCAGTACGCTG 333
 DB 96 ArgIleAlaIleValArgAspGlyIleGlnSerAlaAspGlyTyrGlyAlaAlaAspIle 115
 QY 334 ACTCCC-----AACGGTTTTCTGCCATGAATACCGGGTTCATGCCCTCTG 381
 DB 116 AsnAspLysTyrGlyArgAsnThrPheSerLeuSerAsn----- 128
 QY 382 GCGCGCATGAGCGTATTGAGGTATTACAGGGGCGCATGTCACACTGTATGCTCTGAT 441
 DB 129 -----ValLysGlnIleGlnValValLysGlyAlaSerSerThrLeuTyrGlySerGly 146
 QY 442 GCGATGGCGGTGTGTGTAATATCATACAGAAAGAAATGCAGACAAATGGCTCTCTCC 501
 DB 147 AlaIleGlyGlyValVal-----IleIleGluSerLysAlaProGluAspTyrLeuTyrHis 165
 QY 502 -----GTCAATGCGGGCTGAATCTCCAGAA---AGCAACAATGG---GGT 543
 DB 166 ArgAspTyrTyrValAspAlaAlaLeuThrTyrSerGlyIleSerAsnArgTyrGlnGly 185
 QY 544 AACAGCAGC-----CAGTTTAATTTCTGG 567
 DB 186 AsnHisAlaLeuAlaMetArgHisGlyAspGlyGluAlaLeuLeuThrIleAspTyrTyr 205
 QY 568 AGCAGTGGTCCCTTGTGGATGATTCTGTCAGCGCTCAGGTACCGGTAGCACACACAG 627
 DB 205 ----- 205
 QY 628 CGTCAGGGTTTCATCGGTCACATCACCTGAGCGATACAGCAGGACGCGTATTCCTTATCCC 687
 DB 206 ---GlnGlyGluGluThrArgAsnPheAsnGlnAspLeu-----TyrAsn 219
 QY 688 ACGAGTCAACAGAAATATATCTTGTGTCACGCTTTCAGCTGGAAGCGGTGCGAGCAGAT 747
 DB 220 ArgGluValAspGlyTyrAsnLeuGlyPheSerHisTyr----- 233
 QY 748 GTGCTCTGGTTTGTAT-----ATGGATACCACCGCGCAGCGTATATGATAAC--- 792
 DB 234 -----TrpLeuAsnAspAlaLeuArgLeuLysThrHisLeuGluTyrPheAspAspTyr 251

QY 793 -----CGGATGGCAACTGGG-----AGTCTGACGGG 822
 Db 252 AlaLysArgGluGlyThrSerSerIleGlnLysAspLysTrpAspLeuValSer 271
 QY 823 GGATATGACCGGACCTCGCTATGAGCAACAA-----ATTTCAGCTGGCTAT 873
 Db 272 PheTyrGlu-----TyrGlnArgSerGlnThrArgLeuAlaSerValGlyAla 287
 QY 874 GATCATATTTCACCTTCGGAACATCGAATCGTATCTGAAC-----TGGAAC 921
 Db 288 AspTyrThrAlaAsnLeu-----SerTrpMetAspThrLeuGluGlyLysPheTyrTrpArg 306
 QY 922 GAGACAGAAATAAAGGTCGTAGCTGTACGCGAGTGTACTGAAGCGGCACAAATGGGG 981
 Db 307 SerThrGluAsnIleThrGlnThrAsnArgLeuMetAlaAsnAspArgSerGlyAlaGly 326
 QY 982 CTTGCGCGTCAGCGCGGAGCTTAAGGAATCGAACCTT-----1020
 Db 327 IleLeuSerTyrArgArgGluLeuArgAspGluGlyPheAsnAspGluAlaLeuGlyAla 346
 QY 1021 ATCCTCAATTTCATTCTGCTTACCTCCTCGGAGAA-----TCTCATCTG 1065
 Db 347 ThrLeuAsnIleGlnLysGluTrpGlnGlnGlyGluTrpLeuHisGlnPheAlaTyrGly 366
 QY 1066 GTTACGCTGGGGGC---GAGTTTCAG-----AGCTCGTCC 1098
 Db 367 MetSerValAspGlyHisAspTyrGlnArgProLysSerIleArgArgMetGluSerSer 386
 QY 1099 ATGAAGACGAGTCTCTCTCCACACAGGTGAACCTTC-----CGGAGAAA 1149
 Db 387 GlyAspAsp-----LeuGlnAlaAspGluProPheAlaProAlaArgGluTyr 402
 QY 1150 AGCTGCTCGGTATTTCTGAGGATGAGTGGCATCTCCAGCATGCTGCTGCTGCTGCTG 1209
 Db 403 ArgPheGlyValTyrGlyGlnAspAsnLeuLeuGlyAsp---TrpThrLeuAlaAla 421
 QY 1210 GGACGCGCTATGAA-----CATCATGACCAATCGGGGACAC 1248
 Db 422 GlyLeuArgPheAspAlaGlnLysLeuSerProLysAsnThrAspArgIleHisGlyTyr 441
 QY 1249 -----TTTCAGTCGCGTGCATATCTGCTCGGATG 1281
 Db 442 LysValValThrMetGlySerSerGluTrpSerProSerAlaSerIleSerTyrGlnTrp 461
 QY 1282 GCAGATGCTGACGCTGAAAGCGGTGTGACCGGATATACCGACCCAGAAATGGG 1341
 Db 462 HisProGluTrpAsnThrTyrLeuSerTyrAsnHisGlyPheArgAlaProSerTyrAsp 481
 QY 1342 CAGCTACATAAAGGATTAGTGTGTCTCGGCGAGGAAACAAATCTACTGGTAAC 1401
 Db 482 LysAlaTyr---GlyAlaSerAspHisSerPheValProLeuThrProPheIleIleLys 500
 QY 1402 CCGAC-----CTGAAGCGGAAGAGCGTCAGTTATGAGGCTGGGTGTTATACAT 1455
 Db 501 ProAsnAsnLysLeuAlaGluThrSerAspSerPheGluLeuGlySerLysTyrAsp 520
 QY 1456 AACCCCGCGTCTGATCCCATGTCACAGGTTTATGACTGACTTCTCCAC-----1509
 Db 521 Asn---GlyGlnThrGlnPheTyrValAlaValPheTyrSerIlePheAspAsnPheIle 539
 QY 1510 -----AAGATTGCTCTCTTATCCATAATGATAACAC-----AATAGC 1548
 Db 540 AspValLysGlnValGlyTyr-----AspAsnAlaThrGlySerValIleGlnGln 556
 QY 1549 TATGTAAACAGCGGAAGCGCGTGTGACGCTGTGGAATTTGCCGCGACATGCGCGTGT 1608
 Db 557 TyrGlnAsnIleAlaGlyValLysThrTyrGlyAlaGluMetSerValMetHisArgLeu 576
 QY 1609 -----TGCTCAGAGGATGTCAGCTGTCTCACTGAATTTACACCTGACCGCAAGTAA 1659
 Db 577 AspAspArgTrpSer-----ValGluAsnLysLeuGlyTyrVal-----Asp 590
 QY 1660 CAACGTGATGGTGATAACAA-----GGTGGCGCG 1689

Db 591 GlyLysAspGlyGluAsnGlnTyrValArgThrLeuThrProLeuGluGlySerValGln 610
 QY 1690 CTGAGTTATACCTCGTAACACATG---GTGAATCGGAAACTGAACTGGCAGATCACCGAA 1746
 Db 611 LeuAsnTyrGlnArgGluArgTrpAspAlaTyrSerArgLeuAsnTrp-----626
 QY 1747 GAGTGGCATCATGGTGGTGGCGCTTATCGCGGAAACACACCGTTTTCACCCAGAAT 1806
 Db 627 -----AlaSerAlaMetSerArgValProThrCysThr-----637
 QY 1807 TATTGCTCACTGACGCGTGTACAGAAAGAGTGTATCATGAGAAAGAGAAATACCTGAAA 1866
 Db 638 -----ThrGluGlnGlyLysGluThrGlu 645
 QY 1867 -----GCCTGACGGTGGTGGATGCGAGTCTGCTGCTGGAAGATGACGGAT 1911
 Db 646 CysAlaThrThrThrGlyTrpValSerTrpAspIleGlyLeuAsnTyrGlnTrpAsnAla 665
 QY 1912 GCCCTGACGCTGAATGCTGCGTGAATAACCTCTCAACAAAGGATTAACAGT-----1962
 Db 666 GlnLeuSerAlaSerPheAsnValValAsnLeuLeuAspArgGluTyrThrArgTyrGln 685
 QY 1963 GAGTGAAGCTGTACAGTGGCGGT-----AAGAGTACGCTGTATGCC-----2004
 Db 686 AspVal-----AlaGlyValThrProSerAspThrLeuTyrSerThrGluPro 701
 QY 2005 GGTGATTACTTC 2016
 Db 702 GlyArgTyrPhe 705
 RESULT 9
 US-10-282-122A-48409
 ; Sequence 48409, Application US/10282122A
 ; Publication No. US20040029129A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Liangsu
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Malone, Cheryl
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari
 ; APPLICANT: Zyskind, Judith
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John
 ; APPLICANT: Carr, Grant
 ; APPLICANT: Yamamoto, Robert
 ; APPLICANT: Forsyth, R.
 ; APPLICANT: Xu, H.
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
 ; FILE REFERENCE: ELITEA.034A
 ; CURRENT APPLICATION NUMBER: US/10/282,122A
 ; CURRENT FILING DATE: 2003-02-20
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/230,335
 ; PRIOR FILING DATE: 2000-09-06
 ; PRIOR APPLICATION NUMBER: 60/230,347
 ; PRIOR FILING DATE: 2000-09-09
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/267,636
 ; PRIOR FILING DATE: 2001-02-09
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; Remaining Prior Application data removed - See File Wrapper or PALM.

```

; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 48409
; LENGTH: 775
; TYPE: PRT
; ORGANISM: Bacteroides fragilis
US-10-282-122A-48409

Alignment Scores:
Pred. No.: 1,47e-19 Length: 775
Score: 321.00 Matches: 155
Percent Similarity: 39.32% Conservative: 121
Best Local Similarity: 22.08% Mismatches: 278
Query Match: 8.60% Indels: 148
DB: 12 Gaps: 30

US-10-625-972-4 (1-2091) x US-10-282-122A-48409 (1-775)
QY 82 ATGATTGTCGGCATCCGGGTATGAGAAAAGCTGACTAACGCAGCCGCCAGCTGTTCT 141
Db 101 ValValValThrGlySerArgThrAlaArgProIleLysMetSerProValThrGln 120
QY 142 GTGATTAGCCAGGAGGAATTGCGATCCAGCCAGGTACCAGATCTGGCGGAGCTCTG --- 198
Db 121 ValLeuGlyGlyLeuAlaLeuValAspAlaGlyTyrSerAsnLeuGlnAlaLeuGln 140
QY 199 AGATCAGTAGAGGGTGTGGATGTTGAAAGTGGTACGGGTAAACCCGA ---GGCTGGAA 255
Db 141 GlnGluThrProGlyLeuAsnIleGln-----LysValGlyPheGlyAsnGlu 156
QY 256 ATCAGCATCCGAGGAATCCAGCCAGGTTCACGCTGATATCTGATTGATGGTGTCTGTCAG 315
Db 157 IleSerMetGlnGlyLeuAspAlaArgHisValLeuPheLeuMetAspGlyGluArgMet 176
QY 316 GCGGAGAGCAGTGCAGTGACTCCCAAC---GGTTTTCTGCCATGAATACCGGTTTCATG 372
Db 177 ThrGly-----AspMetAlaGlyAsnLeuAspTyrGluArgPheAsn----- 190
QY 373 CCCCTCTCCGCCCATTCAGCGTATTGAGGTTATCAGGTTATCAGGGGGCCGATGCCACACTGTAT 432
Db 191 -----LeuHisAlaIleAspArgValGluIleValLysGlyAlaSerThrLeuTyr 208
QY 433 GGCTCTGATCGATGGCGGGTGTGTGTAATCATCTACCAGAAGAATGCAGACRAATGG 492
Db 209 GlySerArgAlaAlaGlyAlaValIleAsnLeuIleThrLysLys---ThrAspLysPro 227
QY 493 CTCTCTCCGTCAATGCAGGGCTGAATCTGCAGGAAGCAACAATGGGTAAACACAGC 552
Db 228 Leu---SerIleAspAlaGlyIle-----ArgTyrGlyGlnMetAsn 240
QY 553 CAGTTTAATTTCTGGAGCAGTGGTCCC-----CTGTGGATGAGTATCTGTCTC 597
Db 241 GluArgAsnTyrLysHisProGlnProLysAspPheLeuTyrMetPheGluGlnAsnAla 260
QY 598 AGCCTCAGGTACCGGTAGCACACACAGCGTCAGGGTTTCATCGGTACATCATCTGAGC 657
Db 261 Asp-----ArgProAsnLeuGlnSerTrpValSerAlaGlyPheLysAlaGlyLys 277
QY 658 GATACAGCAGCACCGGTATTCCTTATCCCGAGGAGTCACAGAATTATATCTTGGTGA 717
Db 278 PheThrSerGlnThrAspValTyrSerGluSerAspAlaPheTyrMetTyrGlnAla 297
QY 718 CGTCTTTCAGGAGCGTCGAGCAGAGATGTGCTCTGTTT----- 759
Db 298 GluAsnAspLysLysValTyrThrLysGluAlaAsnProPheLeuProHisAspIlelle 317
QY 760 -----GATATGCGATACC 771
Db 318 ValValSerAsnAlaValArgProProMetGlyIleGluGlyLysGluHisIleThrVal 337
QY 772 ACCCGGCAGCGTTTATGATAACCGGGATGGCAACTCGGGAGCTCTGACGGGGGA ----- 825
Db 338 SerGlnLysLeuTyrTyrAsnProAsnProAsnLeuSerValLeuValTyrGlySerSer 357

```


172 QY CAGTACCAGCATCGCGGAGGCTCTGAGATCAGTAGAGGCTGGATGTTGAAGTGGT 231
173 DB |||
95 DB GlnPheGlnSerValLysGluAlaLeu----- 103
232 QY ACCGGTAAACCGGAGGCTGGAATCAGCATCCGAGGAATGCCACCGAGTTACACGGTG 291
233 DB |||
104 DB --GlyTyrThrProGlyValThrValSerSerArgGlyAlaSerAsnThrTyrAspPhe 122
292 QY ATACTGATTGATGGTGTCTCGAGCGGAGGAGCAGTGCAGTACTCCCAACGGTTTT--- 348
293 DB |||
123 DB ValIleIleArgGlyPheSerSerValGlyLeuSerGln-----AsnAsnTyrLeu 139
349 QY -----TCTGCCATGAATACCGGGTTCATGCCCTCTGCGCGCC 387
140 DB |||
140 DB AspGlyLeuLysLeuGlnGlyAsnPheTyrAsnAspAlaValIleAspProTyr---Met 158
388 QY ATTGACGGTATTGAGTTATCAGGGGGCGGATGTCCACACTGATGCTCTGATCGATG 447
389 DB |||
159 DB LeuGluArgValGluLeuMetArgGlyProThrSerValLeuTyrGlyLysSerAsnPro 178
448 QY GCGGCTGTGTGAATATCATTACAGAAAGAAAGTGCAGACAAATGGCTCTCTCCGTCAT 507
449 DB |||
179 DB GlyGlyIleIleSerMetValSerLysArgProThrThrGluProLeuLysGluIleGln 198
508 QY GCAGGGCTGAATCTGCAGGAAACAAACAAATGGGTAAACAGCAGCGAGTTTAAATTCCTGG 567
199 DB |||
199 DB PheLysMetGlyThrAspAsnLeuPheGlnThrGly-----PheAspPheSer 214
568 QY AGCAGTGGTCCCTTCTGGATGATCTGTCTGACGCTCAGCCTCCAGTACGC----- 612
215 DB |||
613 QY GGTAGCACACACAGCGTCCAGGCTTCATCGGTTCACATCATCTGAGCGATACAGCAGCCAG 672
232 DB |||
673 QY CGTATTCCTTATCCCGAGGAGTCAAGAAATTAATCTTGGTGCAGCTTGCAGTGGAG 732
242 DB |||
733 QY GCGTCGAGCAGGATGTCTCTGTTT-----GATATGGATACCAACC 774
256 DB |||
775 QY CGGACGGTTATGATACCCGGAGTGGCAACTGGGGAGTCTGACGGGGGATATGACCGG 834
276 DB |||
835 QY ACCCTG-----CGCTATGACGGAACAAATTT 861
294 DB |||
862 QY TCAGCTGGCTAT-----GATCATACTTTC-----ACCTTCGGAACATGGAATCGTAT 909
314 DB |||
910 QY CTGAAGTGAACGACAGACAGAAATAAAGTCTGAGCTTGTACGCGAGTGTACTGAGCGC 969
333 DB |||
970 QY GACAAATGGGGCTTCCCGTTCAGCGCGGAGCTTAAGGAATCGAACCTTATCTCGAAT 1029
349 DB |||
1030 QY TCATTACTGCTTACC-----CCTCTGGGAGAA 1056
366 DB |||
1057 QY ---TCTCATCTGGTTCGGTGGGGCGAGTTTCAGAGCTCGTCCATGAAAGACGGAGTT 1113
386 DB |||
1114 QY GTCCTTGCAGCAC----- 1128

406 DB ThrPheGlySerAlaProSerIleAspLeuTyrAsnAsnTyrHisProGluTyrPheAla 425
1129 QY -----GGTGAACCTTTCCGG-----CAGAAAAGCTGGTATTTGGTCTGAG 1170
426 DB |||
426 DB PheGlyGlyAlaGluProTyrGlnMetAsnLysSerGlnThrGlyLeuTyrValGln 445
1171 QY GAT-----GAGTGGCATCTCAGCGATGCTCGCTGCTGCTGCGGACCGCGCTATGAA 1224
446 DB |||
446 DB AspGlnAlaGluTrp-----AsnLysTrpValPheThrLeuGlyArgTyrAsp 462
1225 QY CATCATGAGCAATTCGGGGGACACTTCAGTCCGGTGCATATCTGCTCGGATGTGGCA 1284
463 DB |||
463 DB TrpSerLysGln----- 466
1285 QY GATCCCTGGACGCTGAAGCGGTGTGACCAAGGATATAAGCACCCAGAAATGGGGCAG 1344
467 DB |||
467 DB --AlaThrThrValArgGlnAsnSerThrThr-----ProThrGluGlyTyr 481
1345 QY CTACATATA-----GGGATTAGTGGTGTCCGGCGCAG 1377
482 DB |||
482 DB IleGluArgAsnAspHisGlnPheThrTrpArgGlyGlyValAsnTyrValPheAspAsn 501
1378 QY GGAATAACA-----AATCTACTT 1395
502 DB |||
502 DB GlyIleSerProTyrPheSerTyrSerGlnSerPheGluProSerAlaPheAspLeuTrp 521
1396 QY GGTAAACCC-----GACCTGAAGCCGGAAGAGAGCGTCAAGTTATGAGCTGGGTGTAT 1449
522 DB |||
522 DB SerThrProArgValSerTyrLysProSerLysGlyGluGlnTyrGluAlaGlyValLys 541
1450 QY TACGATAACCCCGCGCTCTGAATGCCAATGTCCACAGGT-----TTTATGACTGACTTC 1503
542 DB |||
1504 QY TCACACAGATTTCTCTTATTCATTAATGATAACCAAT-----AGCTATGTAAC 1557
561 DB |||
561 DB ThrAsnAsnLeuThrAla-----AspProThrAsnProLeuAlaGlnValPro 576
1558 QY AGCGGAAGCCCGCTTGACGCTGGAAATTTGCCGCGACATTTGCCGCTGGTGGTCAGAG 1617
577 DB |||
577 DB AlaGlyGluIleArgAlaArgGlyValGluLeuGluAlaLysAlaLeu---ThrAla 595
1618 QY GTTCTACGCTGTCACTGAATTAACCTGGACCCGGAAGTGA-----CAACGCTGATGTGAT 1674
596 DB |||
596 DB AsnIleAsnMetThrAlaSerTyrThrThrAspAlaGluTyrThrLysAspThrAsn 615
1675 QY AACAAAGTGGCGCGCTGAGTTATACCTGACATCGTGAATCGCAAACTGAACTGCG 1734
616 DB |||
616 DB LeuLysGlyAsnThrProGluGlnValProGluHisMetAlaSerLeuTrpGlyAspTyr 635
1735 QY CAGATCACCGAAGAGGTGCATCATGGCTG-----GGTCCCGCTTATCCGCG 1782
636 DB |||
636 DB ThrPheAsnGluGlyProLeuSerGlyLeuThrLeuGlyThrGlyArgPheIleGly 655
1783 QY AAACACACCGCTTCCACCGAATATTCTGCTAGCGCTGTACAGAAAGATGTAT 1842
656 DB |||
656 DB SerSer-----Tyr 658
1843 QY GATGAGAAAGGAGATATCCTGAAAGCCTGGACG-----GTGGTGGATCAGGTCTGTGCG 1896
659 DB |||
659 DB GlyAspProAlaAsnSerPheLysValGlySerAlaAlaValMetAspAlaValLys 678
1897 QY TGAAGATGACG-----GATGCCCTGACCTGAATGCTCGGTGAAT 1938
679 DB |||
679 DB TyrAspLeuAlaArgPheGlyMetAlaGlySerSerIleAlaValAsn-----ValAsn 696
1939 QY AACCTGCTCAACAAAGGATTAC 1959
697 DB |||
697 DB AsnLeuLeuAspArgGluTyr 703

RESULT 12

US-10-282-122A-48639

```

/ Sequence 48639, Application US/10282122A
/ Publication No. US20040029129A1
/ GENERAL INFORMATION:
/ APPLICANT: Wang, Liangsu
/ APPLICANT: Zamudio, Carlos
/ APPLICANT: Malone, Cheryl
/ APPLICANT: Haselbeck, Robert
/ APPLICANT: Ohlsen, Kari
/ APPLICANT: Zyskind, Judith
/ APPLICANT: Wall, Daniel
/ APPLICANT: Trawick, John
/ APPLICANT: Carr, Grant
/ APPLICANT: Yamamoto, Robert
/ APPLICANT: Forsyth, R.
/ APPLICANT: Xu, H.
/ TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
/ FILE REFERENCE: ELITRA.034A
/ CURRENT APPLICATION NUMBER: US/10/282,122A
/ CURRENT FILING DATE: 2003-02-20
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/230,335
/ PRIOR FILING DATE: 2000-09-06
/ PRIOR APPLICATION NUMBER: 60/230,347
/ PRIOR FILING DATE: 2000-09-09
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/267,636
/ PRIOR FILING DATE: 2001-02-09
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 78614
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 48639
/ LENGTH: 763
/ TYPE: PRT
/ ORGANISM: Bacteroides fragilis
US-10-282-122A-48639

Alignment Scores:
Pred. No.: 9,298-18 Length: 763
Score: 301.00 Matches: 161
Percent Similarity: 36.94% Conservative: 112
Best Local Similarity: 21.79% Mismatches: 334
Query Match: 8.07% Indels: 132
DB: 12 Gaps: 28

US-10-625-972-4 (1-2091) x US-10-282-122A-48639 (1-763)
Qy 13 ACTCTGGCTCCGTAGTCAATCCCTGCTCCGATTTTCAGCCAGCACCATAGCTGCTGCA 72
Db 91 ThrValSerAlaGluAlaThrSerValValHisPheGlnMetGluGluValSerPheThr 110
Qy 73 GAGAGTGTGATGATGCTCCGCATCCGGCTATGAGAAAAGCTGACTAACGACGCGGCC 132
Db 111 ThrAspGluValValSerAlaAsnArgAsnGluValSerArgLysAlaAlaProVal 130
Qy 133 AGTGTCTTCTGTGATGAGCAGGAGGAATTGAGTTCGAACTGGTACCGGTAAACCGGAGGGCTG 252
Db 131 ValValAsnValMetSerAlaLysLeuPheGluThrValAsnSerThrAspLeuAlaLys 150
Qy 193 GCTCTGAGTCACTAGACGGTGTGATGTTGAACTGGTACCGGTAAACCGGAGGGCTG 252
Db 151 SerLeuAsnPheGlnSerGlyLeuArgValGluAsnAsnCysGlnAsnCysGlyPhePro 170

```

```

QY 1249 ---TTCAGTCGGGTGCATATCTGCTGGATGTGCAGATGCCTGGACGCTGAAGGC 1305
Db 484 IlePheSerProArgValAsnPheLeuTyzLysProSerAspAsnLeuGlnAlaArgLeu 503
QY 1306 GGTGTACACCGGATATAAGGACCCAGCAATGGCGAGCTACATAAAGGGATTAGTGT 1365
Db 504 ThrTySerThrGlyPheArgAlaProGlnAlaTyzAspGluAspLeuHisValThrAla 523
QY 1366 GTGTCCGGCAGGAGGAAAAAACAATCTACTTGGTAACCCGACCTGAAGCCGGAAGAGC 1425
Db 524 ValGlyGlyGluGlyValGlnIleArgLeuAlaAsp---GlyLeuArgGluGlyCysSer 542
QY 1426 GTCAATTATCAGGTCGGGTGTATTACGATAACCCCGCGGT-----CTGAATGCC 1476
Db 543 AsnSerPheSerGlySerValAspTrpSerPheProMetGlyHisTrpGlnSerAsnIle 562
QY 1477 AATGTACAGTTTATGACTACTTCTCCACAAGATTGTCTCTTATTCCATAATGAT 1536
Db 563 LeuLeuGluGlyPheTyzThrAspLeuHisValPheValLeuGluAspIleGlyGlu 582
QY 1537 AACACCAAT-----AGCTATGTAACAGCGGAAGCCCGGTGCACGGT 1581
Db 583 AspGlnAsnGlyAspLysIleLysGluArgArgAsnGlySerGlyAlaLysValTyzGly 602
QY 1582 GTGGAATTTCCGGGCACATTCGCGCTGTGTGCAGAGGATGTCACGCTGTCACCTGAATAC 1641
Db 603 ValAsnLeuAspAlaLysVal---AlaHisGlyArgGluAlaGlnLeuGluPhe 621
QY 1642 ACC-----TGGACCGAAGTGAACAACAGTGTAT 1668
Db 622 ThrValGlnArgSerArgTyzAsnArgAlaGluValTrpSerGluGlyGluGlu 641
QY 1669 GGTGATAACAAGGTGCGCGCTGAGTTATACCCCTGAACAC-----ATG 1713
Db 642 GlnThrThrLysArgMetPro-----ArgThrProAspTyzTyzGlyTyzPheThrPhe 659
QY 1714 GTGAATCGGAATCG---AAGTGGAGATCACCGAAGAGGTGGCATCATGGCTGGTGCC 1770
Db 660 ThrSerAlaProLeuLysAsnPheAspPheSer-----LeuSerGly 673
QY 1771 CGTTATCGGGGAAAAACACACAGTTTTCACCCAGATTATTCTCATGAGCGCTGTACAG 1830
Db 674 ThrTyzThrGlyLys-----MetIleValProHis 683
QY 1831 AGAAAGTGTATGATCAGAAAGGAGAAATACCTGAAGCCTGGACGCTGTGTGGATCAGGT 1890
Db 684 MetAlaGlyTyzIleGluLysSerArgMetGluHisThrProGlnPheMetAspLeuAsn 703
QY 1891 CTGTCTGGAGATGACGATGCCCTG-----ACGCTGATCTCGG 1932
Db 704 LeuLysLeuAsnTyzThrPheValLeuLysAspHisIleLysMetGlnValAsnGlyGly 723
QY 1933 GTGAATAACCTGTCTCAACAGGATTACAGTGACGTGAGCGCTGTACAGTGGCGGTAAAGT 1992
Db 724 ValGlnAsnIlePheAsnSerPheGlnLysAspLeuAsp-----LysGly 738
QY 1993 ACCTGTATCGCGGTGATTACTTCCAGACGGGATCATCAACACAGGATATGTGATA 2049
Db 739 GluPheArgAspAlaGlyTyzPheTyzGlyProThrGlnProArgThrTyzPheVal 757

```

RESULT 13

US-10-282-122A-43232
 ; Sequence 43232, Application US/10282122A

; Publication No. US20040029129A1

; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu

; APPLICANT: Zamudio, Carlos

; APPLICANT: Malone, Cheryl

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari

; APPLICANT: Zyskind, Judith

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John

```

; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITEA 034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 43232
; LENGTH: 774
; TYPE: PRF
; ORGANISM: Escherichia coli
US-10-282-122A-43232

```

Alignment Scores:

```

Pred. No.: 1,04e-17 Length: 774
Score: 300.50 Matches: 169
Percent Similarity: 37.52% Conservative: 109
Best Local Similarity: 22.81% Mismatches: 282
Query Match: 8.05% Indels: 181
DB: 12 Gaps: 40

```

US-10-625-972-4 (1-2091) x US-10-282-122A-43232 (1-774)

```

QY 64 GCTGCTGCAGAGATGTGATGATGTC-----TCGGCATCCGCTATGAGAA 111
Db 109 AlapProLysGluAspAlaLeuThrValValGlyAspTrpLeuGlyAspAlaArgGluAsn 128
QY 112 AAGCTGACTAACGACGCGCCAGTGTCTGTGATTAGCCAGGAGAAATTCAGTCCAGC 171
Db 129 AspValPheGluHisAlaGlyAlaArgAspValIleArgArgGluAspPheAlaLysThr 148
QY 172 CAGTACACAGATCGCGGAGGCTCTGAGATCAGTAGAGGCTGTGGATGTT-----GAA 225
Db 149 GlyAlaThrThrMetArgGluValLeuAsnArgIleProGlyValSerAlaProGluAsn 168
QY 226 AGTGGTACGGTAAAAACCGGA---GGCTGGAATTCAGCATCCGAGGAATGCCA----- 276
Db 169 AsnGlyThrGlySerHisAspLeuAlaMetAsnPheGlyIleArgGlyLeuAsnProArg 188
QY 277 ---GCCAGTTACAGCTGATGATGATGATGTTGTCAGGGCGGGAAGCAGTACGCTG 333
Db 189 LeuAlaSerArgSerThrValLeuMetAspGlyIle-----ProValProPhe 204
QY 334 ACTCCCAACGGTTTTCTGCCATGATACCGGTTTCATCCCCCT-----CTGCCCGCC 387
Db 205 AlaProTyzGlyGlnProGlnLeuSer-----LeuAlaProValSerLeuGlyAsn 221
QY 388 ATTACGCTATTGAGGTATTATCAGGGCGCGCATGTGCACACTGTATGGTCTGTGATCGATG 447
Db 388 ATTACGCTATTGAGGTATTATCAGGGCGCGCATGTGCACACTGTATGGTCTGTGATCGATG 447

```

Db 222 MetAspAlaIleAspValValArgGlyGlyAlaValArgTyrGlyProGlnSerVal 241
 Qy 448 GCGCGTGTGGTGAATATCATACCAGAGAAATGACAGAAATGCGCTCTTCGGTCAAT 507
 Db 242 GlyGlyValValAsnPheValThrArgAlaIleProGlnAspPhe-----GlyIleGlu 259
 Qy 508 CGAGCGTGAATCTGCAGGAAGCAACAAATGGGGTAACAGCAGCGATTTAATTTCTGG 567
 Db 260 AlaGlyValGluGlyGlnLeuSer-----ProThrSerSerGlnAsnAsn----- 274
 Qy 568 AGCAGTGGTCCCTTGTGGATGATTCTGTACCTGCGAGTACCGGTGACACACAACAG 627
 Db 275 -----ProLysGluThrHisAsnLeuMetValGlyGlyThrAlaAspAsn 289
 Qy 628 CGTACGGTTTCATCGTCACTACCTAGCAGATACAGCAGCAGCGGTATCTTATCCC 687
 Db 290 GlyPheGlyThrAlaLeuLeu---TyrSerGlyThrArgGlySer----- 303
 Qy 688 ACGGAGTCACAGAAATTAATCTTGGTCACGCTCTGTACTGGAAG-----CGGTG 738
 Db 304 -----AspTyrArgGluHisSerAlaThr 311
 Qy 739 GAGCAGGATGCTCTGCTTT-----GATATGAT 768
 Db 312 ArgIleAspAspLeuMetLeuLysSerLysTyrAlaProAspGluValHisThrPheAsn 331
 Qy 769 ACCACCGCGCGGTATGATAACCGGATGGCAACTGGGAGTCTG----- 816
 Db 332 SerLeuLeuGlnTyrTyrAspGlyAlaAspMetProGlyGlyLeuSerArgAlaAsp 351
 Qy 817 -----ACGGGGGATATGACCGGACCGCTCGCTGATGAGCA 852
 Db 352 TyrAspAlaAspArgTyrGlnSerThrArgProTyrAspArgPhe-----TyrArg 369
 Qy 853 AACAAATTT-----TCAGCTGGCTAT-----GATCATCTTTCACCTTC 891
 Db 370 ArgLysLeuAlaSerLeuGlyTyrGlnPheGlnProAspSerGlnHisLysPheAsnIle 389
 Qy 892 ---GNAACATGAAATCGTATCTGAACCTGGAACGAGACAGAAATAAAGTCTGAGCTT 948
 Db 390 GlnGlyPheTyrThrGlnThrLeuArgSerGlyTyrLeuGlu---GlnGlyLysArgile 408
 Qy 949 GTACCGAGTACTGAGCGGACAAATGGGGGCTTGGCGGTACCGCGGAGCTTAAG 1008
 Db 409 ThrLeuSer-----ProArgAsnTyrTrpValArgGlyIleGluProArg----- 423
 Qy 1009 GAATCGAACCTTATCTGAATTCATTACTGCTTACCCCTCTGGGAGATCT---CATCTG 1065
 Db 424 TyrSerGlnIlePheMet-----IleGlyProSerAlaHisGlu 436
 Qy 1066 GTTACGGTGGGGCGAGTTTCAGAGCTCGTCCATGAAGACGAGGTGCTTCCGCCAGC 1125
 Db 437 ValGlyValGlyTyrArgTyrLeuAsnGluSerThrHisGluMetArgTyrThrAla 456
 Qy 1126 ACAGGTGAACCTTCCGGCAGAAAGCTGGTGGTATTGCTGAGAN----- 1173
 Db 457 ThrSerSerGlyGlnLeuProSerGlySerProTyrAspArgAspThrArgSerGly 476
 Qy 1174 -----GAGTGGCATCTCAGGATGCACTT-----CGCGTGAAT 1206
 Db 477 ThrGluAlaHisAlaTyrTyrLeuAspAspLysIleAspIleGlyAsnTrpThrIleThr 496
 Qy 1207 GCGGGCAGCGCTATGAATCATATGAGCAATTC-----CGGGACAC 1248
 Db 497 ProGlyMetArgPheGluHisIleGluSerTyrGlnAsnAsnAlaIleThrGlyThrHis 516
 Qy 1249 TTCAGT-----CCGGTGCATATCTGTGCTGGGATGTCGAGAT 1287
 Db 517 GluGluValSerTyrAsnAlaProLeuProAlaLeuAsnValLeuTyrHisLeuThrAsp 536
 Qy 1288 GCCTGGACGCTGAAGCGCGGTGTACCCAGCGGATATAGCCACCCAGAAATGGGCGACTA 1347
 Db 537 SerTyrAsnLeuTyrAlaAsnThrGluGlySerPheGlyThrValGlnTyrSerGlnIle 556

RESULT 14

US-10-282-122A-55554
 ; Sequence 55554, Application US/10282122A
 ; Publication No. US20040029129A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Liangsu
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Malone, Cheryl
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari
 ; APPLICANT: Zyskind, Judith
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Carr, Grant
 ; APPLICANT: Yamamoto, Robert
 ; APPLICANT: Forsyth, R.
 ; APPLICANT: Xu, H.
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
 ; FILE REFERENCE: ELITRA.034A
 ; CURRENT APPLICATION NUMBER: US/10/282,122A

Db 500 ProTyrPheSer-----TyrSerGluSerPheGluProAlaSerGlnThrAspAla 516
QY 1375 CAGGAAAAAACAATCTACTTGTGTAAACCCGACCTGAAGCCGGAAGAGCGTCAGTTAT 1434
Db 517 GlnGlyLysLeu-----PheSerProSerLysGlyLysGlnTyr 529
QY 1435 GAGCTGGGTGTATTAC-----GATAACCCCGCGGTCTGAATGCCAATCTCACA 1485
Db 530 GluAlaGlyValLysTyrValProAsnAspArgPro-----IleValValThr 545
QY 1486 GGT-----TTATGACTGACTCTCCAAAGATTTCTCTTATTCATAATGATAAC 1539
Db 546 GlyAlaLeuTyrGlnLeuThrLysThrAsnLeuMetAla-----AspPro 561
QY 1540 ACCAATAGCTAT-----GTAACACGCGAAAGCCGTTCCACGGTGTGGAATTTGCC 1593
Db 562 AlaGlySerPhePheSerValGlnGlyGlyGluIleArgAlaArgGlyValGluLeuGlu 581
QY 1594 GGCACATTGGCTGGTGCAGAGGATGTCACGCTGCTCACTGAATACACCTGGACCCGA 1653
Db 582 AlaLysAlaAlaLeu---SerAlaSerValAsnLeuValGlySerTyrThrThrAsp 600
QY 1654 AGTGAA---CAACGTGATGTGTATGAACAAAGTGGCCGCTGAGTTATACCCCTGAACAC 1710
Db 601 AlaGluTyrThrThrAspThrThrTyrLysGlyAsnThrProAlaGlnValProLysHis 620
QY 1711 ATGTGTAATCGGAACCTGAACCTGCAGATCACCGAAGGTGGCATCTGGCTG----- 1764
Db 621 MetAlaSerLeuTrpGlyAspTyrThrLeuPheAspGlyAlaLeuSerGlyLeuThrLeu 640
QY 1765 -----GGTCCCGTTATCGCGGAAACACACACGCTTCCACCCAGATTAATTTCGCTG 1818
Db 641 GlyThrGlyValArgTyrThrGlySerSer----- 650
QY 1819 AGCGCTGTACAGAAAGTGTATGATGAGAAAGGAGATACCTGAAG-----GCCTGG 1872
Db 651 -----TyrGlyAspProAlaAsnSerPheLysValGlySerTyr 663
QY 1873 ACGGTGGTGATCAGGTCTGTCTGTGAAGATG-----ACGGATGCCCTGACGCTGAAT 1926
Db 664 ThrValValAspAlaLeuValArgTyrAspLeuAlaArgValGlyMetAlaGlySerAsn 683
QY 1927 GCTGCG-----GTGAATACCTGCTCAACAGGATTAC 1959
Db 684 ValAlaLeuHisValAsnAsnLeuPheAspArgGluTyr 696

RESULT 15
US-10-282-122A-69342
; Sequence 69342, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsythe, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 69342
; LENGTH: 755
; TYPE: PRT
; ORGANISM: Pseudomonas syringae
US-10-282-122A-69342

Alignment Scores:
Pred. No.: 4,39e-17 Length: 755
Score: 293.50 Matches: 173
Percent Similarity: 36.71% Conservative: 117
Best Local Similarity: 21.90% Mismatches: 271
Query Match: 7.86% Indels: 229
DB: 12 Gaps: 40

US-10-625-972-4 (1-2091) x US-10-282-122A-69342 (1-755)

QY 4 CGAATAACACACTCTGGCTTCCTAGTCACTCCCTCTCGGATTTTCAGCCAGCAGCAT 63
Db 31 ArgIleGluThrAlaProGluAlaValGluGlySerLeuGluAlaThrThrfile 50
QY 64 GCTGCTCAGAGAGTGTGATGATT----- 87
Db 51 SerGlyAlaSerArgLeuGluThrAlaThrGlyProValThrGlyTyrValAlaThrArg 70
QY 88 ---GTCTCCGGCATCCGGCTATGAGAAAAGTGTACTAAGCAGCGCCGCGTTCGTG 144
Db 71 GlyLeuSerAlaThrLysThrAspThrAlaLeuIleGluThrProGlnSerIleSerVal 90
QY 145 ATTAGCCAGGAGAAATTCAGATCCAGCCAGCAGTACACAGATCTGGCGGAGGCTCTGAGATCA 204
Db 91 ValThrLysAspGlnMetLysAlaGlnGlyAlaGluAsnLeuSerGlnMetLeuArgTyr 110
QY 205 GTAGAGGTGTGGATGTTCAAGTGTACGGGTAAACCGGAGGGCTGAA---ATCAGC 261
Db 111 SerAlaAlaValValProGluThrArgGlySerThrAlaSerArgLeuAspMetLeuSer 130
QY 262 ATCCGAGGAATG---CCAGCCAGTTACACGCTGATCTGATTTGTTGTTCTGTCAGGGC 318
Db 131 IleArgGlyPheSerProAlaLeuTyr-----LeuAspGlyLeuArgMetPro 146
QY 319 GGAACAGTCAGTCAGTCCCAACCGTTTCTGCCATGAATACGGGTTTCATGCCCCCT 378
Db 147 AspAsnArgAspAlaAlaPro-----GlnLysAsp 156
QY 379 CTGGCGCCCATTCAGCGTATTGAGTTATCAGGGGGCGGATGTCCACACTGTATGGTCTCT 438
Db 157 ValPheAspLeuGluArgValGluValLeuArgGlyProAlaSerValLeuTyrGlyGln 176
QY 439 GATGCGATGGCGGTGTGTGAATATCATACCAGAAAGATGCGACAAATGGCTCTCT 498
Db 177 AlaSerProSerGlyValValAsnMetValSerLysLeuProThrGluThrProPheHis 196
QY 499 TCCGTCAATGCGGGCTGAATCTGCAGGAAACCAACAAATGGGTAAACAGCAGCAGTTT 558
Db 197 GluIle-----GlyLeuThrTyrGlyThrPheAsnLys-----LysArgThrThrPhe 212

QY	559	AA	TTTCTGGAGCAGTGCTCCCTTGTGGATGATCTGTC	---AGCTCGAGGTACGCGT	615		
Db	213	Asp	He	:::	:::		
QY	616	AGC	ACACACAGCGTCAGGGTTCATCGGTACATC	ACTGACCGGATACAGACGACGACCGT	675		
Db	231	---	---	:::	:::		
QY	676	ATT	CTCTTCCACGGAGTCACAGAA	TTATAATCTTGGTCACGCTTGACTGGAAGCGC	735		
Db	239	Ile	Glu	His::: :::	:::		
QY	736	TCG	GACGAGGAGTGCTCTGCTGTTTCATATG	TACCAACCGCGCAGCGTTATGATAACCGG	795		
Db	258	Asp	Glu	AspThrSerLeu	ThrLeu::: :::		
QY	796	GAT	GGCAACTGGGAGTCTG	-----Leu	AlaSerLeuGlnSerAspPro	825	
Db	275	Gly	Ala	SerTyrGlySerMetPro	AlaTyrGlySerValValLysSerPro	825	
QY	826	---	---	TATGACCGGACCTCGCTATGAG	-----CGAAACAAA	858	
Db	295	His	Ile	AspPheAspPheTyrAsp	GlyGluLysAsnPheGluLysSerAspArg	915	
QY	859	ATT	T	CAGCTGGCTATGATCATCTACTT	CCGTAACCTCGGAACATCGTAATCTATCTGAAC	915	
Db	315	His	Ala	ValGlyTyr	-----Leu	PheGluHisLeuAsnAsp	915
QY	916	---	TGG	AACGACAGACAGAAAT	-----AAAGTCGTGAGCTTGACGAGTGA	960	
Db	328	Val	Trp	ThrLeuArgGlnAsnAlaArgTyr	LeuArgSerGluGlyValTyrArgSerIle	347	
QY	961	CTG	AAGCGCAGCAAAATGGGCTTGGCGTCAGCG	-----CGGAG	1002		
Db	348	Tyr	-----	AsnGlyTrpGlyThrLeu	---GlnPro	AspTyrArgThrSerGluArgAla	364
QY	1003	CTT	AAGGAATCGAACCTTATCTCTGAATTCAT	TACTGCTTACC	-----	1044	
Db	365	Thr	Ile	AlaThrAspValAsnLeuAspSerTyr	ThrIleAspAsnGlnMetGlnAlaAsn	384	
QY	1045	---	---	CCTCTGGGAGATCTCATCTGGTTACG	TGGGGCGGAGTTTCAGAC	1092	
Db	385	Phe	Asp	ThrGlyProLeu	-----GlnHis	ThrLeuLeuGlyAlaAspTyrGlnAsn	402
QY	1093	TCG	TCC	-----ATGAAAGCGGAGTTGTCTCTGCC	AGCAGAGTGAACCTTTC	1140	
Db	403	Thr	Ser	ThrAspThrLysAlaGlyTyrGly	IleGlyProThrLeuAspIlePheAspPro	422	
QY	1141	---	---	CGG	1143	---	---
Db	423	Val	Tyr	GlySerProValGluValPro	AlaPheThrGluSerSerThrGlnArgAspGln	442	
QY	1144	CAG	AAAAGCTGGTGGTATTTGCTGAGATGAG	TGGCATCTCAGGATGCACCTTGGCTG	1203		
Db	443	Gln	Lys	GlyLeuTyrLeuGlnGluGlnLeu	LysTrp	-----AspLysTrpValLeu	459
QY	1204	ACT	GCGGCGACCGCTATGACATCATGACA	ATTCGCGGACACTTCAGTCGCGTGA	1263		
Db	460	Leu	Met	GlyGlyArgTyr	-----	465	
QY	1264	TAT	CTGCTCGGATGTCGAGATGCTGGACGCT	GAAAGCGGTGTGACCGGATAT	1323		
Db	466	---	---	AspTrp	AlaGluSerSerAsnSerSerThrAsnLeuThrSerThr	481	
QY	1324	AAG	-----	GCACCAGATGGCGGAGCTACATAAA	1353		
Db	482	Lys	Thr	ArgSerSerAlaAspSerGluAlaPhe	ThrGlyArgLeuGlyLeuValTyrLeu	501	
QY	1354	GGG	ATTAGTGGTGTG	-----	TCCGG	1374	
Db	502	Phe	Asp	AsnGlnGlyLeuAlaProTyr	IleSerTyrSerGluSerPheGluProGlnSerGly	521	
QY	1375	CAG	GGAACCAATCTACTTGTGAACCCCGAC	CTCGAAGCGGAGAGAGCGCTCAGTTAT	1434		

Search completed: October 13, 2004, 12:16:35
Job time : 178 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: October 13, 2004, 11:37:12 ; Search time 24.5 Seconds

(without alignments)
8888.057 Million cell updates/sec

Title: US-10-625-972-4

Perfect score: 3732
Sequence: 1 atgcgaataaccactctggc.....cgctgaactatcagttctga 2091

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame_n2p.model -DEV=xlh
-Q/cn2_1/USPTO spool/US10625972/runat 12102004 130105 26806/app query.fasta 1.2247
-DB=SwissProt_42 -QPM=fastan -SUFFIX=isp -MINMATCH=0.1 -LOPCCL=0 -LOCPXT=0
-UNITS=bits -START=1 -END=1 -MATH=blcsum62 -TRANS=human40 cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10625972 @CN 1 1 29 @runat 12102004 130105 26806 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THRADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1052.5	28.2	652	1 IRGA_VIBCH	P27772 vibrio chol
2	832.5	22.3	663	1 IRGA_VIBCH	P17315 escherichia
3	625	16.7	746	1 FEPA_ECOLI	P05825 escherichia
4	592.5	15.9	746	1 FEPA_PSEAE	Q05098 pseudomonas
5	528.5	14.2	614	1 BTUB_ECOLI	P06129 escherichia
6	510.5	13.7	614	1 BTUB_SALTY	P37409 salmonella
7	373.5	10.0	700	1 YNCD_ECOLI	P76115 escherichia
8	355	9.5	687	1 HEME_YEREN	P31499 versinia en
9	341	9.1	676	1 HMUR_YERPE	Q56989 versinia pe
10	339.5	9.1	673	1 FYUA_YERPE	P46359 versinia pe
11	332.5	8.9	673	1 FYUA_YEREN	P46360 versinia en
12	301.5	8.1	760	1 YBIL_ECOLI	P75780 escherichia
13	300.5	8.1	774	1 FECA_ECOLI	P13036 escherichia
14	291.5	7.8	723	1 Y262_HAEIN	P43600 haemophilus
15	291	7.8	725	1 HXC2_HAEIN	P45357 haemophilus
16	282.5	7.6	720	1 FFTA_PSEAE	P42512 pseudomonas
17	280.5	7.5	702	1 FOXA_SALTY	Q56445 salmonella
18	278	7.4	747	1 FHUA_ECOLI	P06971 escherichia

19	266.5	7.1	710	1 FOXA_YEREN	Q01674 versinia en
20	262.5	7.0	743	1 BFED_BORPE	P81549 bordetella
21	260	7.0	819	1 PUPA_PSEPU	P25184 pseudomonas
22	253	6.8	746	1 RHTA_RHIME	Q923Q5 rhizobium m
23	248.5	6.7	726	1 FATA_VIBAN	P11461 vibrio angu
24	244	6.5	810	1 HPUB_NEIMC	P96949 neisseria m
25	241.5	6.5	810	1 HPUB_NEIMA	Q91wa2 neisseria m
26	239	6.4	944	1 LBPA_ECOLI	Q9Jtk4 neisseria m
27	233	6.2	729	1 FHUE_ECOLI	P16869 escherichia
28	231.5	6.2	732	1 IUTA_ECOLI	P14542 escherichia
29	230	6.2	908	1 TB12_NEIMO	Q06987 neisseria g
30	228.5	6.1	915	1 TB11_NEIGO	Q01996 neisseria m
31	225.5	6.0	790	1 YDBB_ECOLI	P31827 escherichia
32	225.5	6.0	911	1 TB11_NEIMB	Q09056 neisseria m
33	220	5.9	735	1 FCT_ERWCH	Q47162 erwinia chr
34	220	5.9	815	1 FPVA_PSEAE	P48632 pseudomonas
35	220	5.9	943	1 LBPA_NEIMB	Q06379 neisseria m
36	214	5.7	345	1 YEGA_HAEIN	Q06241 haemophilus
37	213	5.7	1067	1 HGBB_HAEIN	Q9Kiv1 haemophilus
38	209	5.6	826	1 PBUA_PSESP	Q08017 pseudomonas
39	207	5.5	687	1 VIUA_VIBCH	Q00964 vibrio chol
40	207	5.5	993	1 HGBC_HAEIN	Q9Kiv0 haemophilus
41	205.5	5.5	1084	1 HGP3_HAEIN	P44836 haemophilus
42	198.5	5.3	809	1 PUPB_PSEPU	P38047 pseudomonas
43	196.5	5.3	999	1 HGP2_HAEIN	P44809 haemophilus
44	193.5	5.2	999	1 HGPB_HAEIN	C87296 haemophilus
45	193.5	5.2	1013	1 HGBA_HAEIN	Q9Kiv2 haemophilus

ALIGNMENTS

RESULT 1
IRGA_VIBCH STANDARD; PRT; 652 AA.
AC P27772; Q9KUP0;
DT 01-AUG-1992 (Rel. 23, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Iron-regulated outer membrane virulence protein precursor.
GN IRGA OR VC0475.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=Classical Ogawa 395 / ATCC 39541 / Serotype O1;
RC MEDLINE=91023868; PubMed=1406279;
RA Goldberger M.B., Boyko S.A., Butters J.R., Stoeber J.A.,
Payne S.M., Calderwood S.B.;
RT "Characterization of a Vibrio cholerae virulence factor homologous to
the family of TonB-dependent proteins.";
RL Mol. Microbiol. 6:2407-2418(1992).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=El Tor N16961 / Serotype O1;
RC MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
Ermlolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
McDonald L., Ueberlack T., Fleischmann R.D., Niernan W.C., White O.,
Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
cholerae".
RL Nature 406:477-483(2000).
[3]
RN SEQUENCE OF 1-151 FROM N.A.
RP STRAIN=Classical Ogawa 395 / ATCC 39541 / Serotype O1;
RC MEDLINE=91072235; PubMed=2174861;
RA Goldberger M.B., Boyko S.A., Calderwood S.B.;
RT "Transcriptional regulation by iron of a Vibrio cholerae virulence

RT gene and homology of the gene to the *Escherichia coli* fur system.";
 RL J. Bacteriol. 172:6863-6870 (1990).
 CC -!- FUNCTION: INVOLVED IN THE INITIAL STEP OF IRON UPTAKE BY BINDING
 CC FERRIC VIBRIOBACTIN, AN IRON CHELATING SIDEROPHORE THAT ALLOWS
 CC V. CHOLERAE TO EXTRACT IRON FROM THE ENVIRONMENT.
 CC -!- SUBCELLULAR LOCATION: Outer membrane.
 CC -!- MISCELLANEOUS: REGULATION OF THE IRGA EXPRESSION IS NEGATIVELY
 CC REGULATED AT THE TRANSCRIPTIONAL LEVEL BY IRON.
 CC -!- SIMILARITY: Belongs to the tonB-dependent receptor family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----
 CC EMBL; U72152; AAC44766.1; -
 CC EMBL; AE004134; AAF93648.1; -
 CC PIR; D82317; D82317.
 CC PIR; S25265; S25265.
 CC HSPP; P05825; IFEP.
 CC TIGR; VC0475; -
 CC InterPro; IPR000531; TonB_boxC.
 CC Pfam; PF00593; TonB_dep_Bac; 1.
 CC PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
 CC PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
 CC Virulence; Outer membrane; Iron transport; Transport; TonB box;
 KW Signal; Receptor; Complete proteome.
 FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 26 652 IRON-REGULATED OUTER MEMBRANE VIRULENCE
 FT PROTEIN.
 FT SITE 33 40 TONB_BOX.
 FT SITE 635 652 TONB_C-TERMINAL_BOX.
 FT CONFLICT 294 294 D -> G (IN REF. 1).
 FT CONFLICT 448 448 K -> Q (IN REF. 1).
 FT CONFLICT 502 502 A -> T (IN REF. 1).
 SQ SEQUENCE 652 AA; 71669 MW; A95F82FEC072EC93 CRC64;

Alignment Scores:
 Pred. No.: 3,916-68 Length: 652
 Score: 1052.50 Matches: 258
 Percent Similarity: 51.80% Conservative: 116
 Best Local Similarity: 35.73% Mismatches: 223
 Query Match: 28.20% Indels: 125
 DB: 1 Gaps: 24

US-10-625-972-4 (1-2091) x IRGA_VIBCH (1-652)

QY 40 CTGGGATTTTCAGCCAGCAGCATAGCT-----GCTGCAGAGAGTGTGATGATT 87
 Db 17 LeuMetPheSerAlaSerAlaPheAlaGlnAspAlaThrLysThrAspGluThrMetVal 36
 QY 88 GTCTCGGATCCGGCTATGAGAAAGCTAGTACTACGAGCCGCGAGTGTTCGTGATT 147
 Db 37 ValThrAlaAlaGlyTyrAlaGlnValIleGlnAsnAlaProAlaSerIleSerValIle 56
 QY 148 AGCCAGGAGGAATTGCAATGCGCCAGTACCAGCATCTGGCGAGGCTCTGAGATCAGTA 207
 Db 57 SerArgGluAspLeuSerArgTyrTyrArgAspValThrAspAlaLeuLysSerVal 76
 QY 208 GAGGTGTGGATGTTGAAAGTGGTACCGGTAAACCGGAGGCTGGGAATCAGCATCCGA 267
 Db 77 ProGlyValThrValThrGlyGlyAspThrThr-----AspIleSerIleArg 93
 QY 268 GGAATGCCAGCAGTTACACGCTGATCTGATGATGTTGTTCGTCAGCGCGGAGCAGT 327
 Db 94 GlyMetGlySerAsnTyrThrLeuIleLeuValAspGlyLysArgGln---ThrSerArg 112
 QY 328 GACGTGACTCCCAACGGTTFCT---GCCATGAATACCGGGTTCATGCCCTCTGGCC 384
 Db 113 GlnThrArgProAsnSerAspGlyProGlyIleGlnGlnGlyTyrLeuProLeuGln 132

QY 385 GCCATTGAGCGTATTGAGTTATCAGGGGCGCGATGATCCACACTGTATGCTGTGATGCG 444
 Db 133 AlaIleGluArgIleGluValIleArgGlyProMetSerThrLeuTyrGlySerAspAla 152
 QY 445 ATGGCGGTGTGGTGAATATCATTACCAGAAAGATGCAGACAAATGGCTCTCTCCGTC 504
 Db 153 IleGlyGlyValIleAsnIleThrArgLysAspGlnGlnGlnTyrSerGlyAsnVal 172
 QY 505 AATGCAGGCGTGAATCTGCAGGAAGCAACAAATGGGGTAACACAGCCAGTTAATTTC 564
 Db 173 GlnLeuSerThrValValGlnGluAsnArgAlaSerGlyAspGluGlnSerAlaAspHe 192
 QY 565 TGGAGCAGTGTCCCTTGTGGATGTTCTGCAGCTGCAGGTACGCGGTAGCACAA 624
 Db 193 PheValThrGlyProLeu---SerAspAlaLeuSerLeuGlnValTyrGlyGlnThr 211
 QY 625 CAGCGTCAG-----GGTTCATCGTCACTCACTGAGC 657
 Db 212 GlnArgAspGluAspGluIleGluHisGlyTyrGlyAspLysSerLeuArgSerLeu 230
 QY 658 GATACAGCAGGCGCGCTATTCTTATCCACCGAGTCACAGAAATTATAATCTTGGT 714
 Db 231 -----ThrSerLysLeuAsnTyrGlnLeuAsnPro 240
 QY 715 -----GCACGCTTGTGAGCGCGCTCGAGCAGGATGTGCTCGTGTGATGATG 765
 Db 241 AspHisGlnLeuGlnLeuGluAlaGlyValSerAlaGlnAsp-----254
 QY 766 GATACACCGCGGAGCGCTTATGATTAACCGGATGGGCACTGGGAGTCTGACGGGGGA 825
 Db 255 -----ArgGluAsnAsnValGlySerAlaGlnSerSerGlyCysArgGlyThr 271
 QY 826 TATGACCGGACC---CTGCGCTATGAGCGAAACAAATTTCACTGCTGATGATGACT 882
 Db 272 CysSerAsnThrAspAsnGlnTyrArgArgAsnHisValAlaValSerHisGln-----289
 QY 883 TTCACCTTCGGAACATCGAAA-----TCGTATCTGAACTGGAACGAG 924
 Db 290 -----GlyAspTrpGlnAspValGlyGlnSerAspThrTyrLeuGlnTyrGluGlu 306
 QY 925 ACAGAAATAAAGTCTGAGCTTGTACGAGTGTACTGAAGCGCAGCAAAATGGGGCTT 984
 Db 307 AsnThrAsnLysSer-----311
 QY 985 GCGGTCAGCGCGGAGCTTAAGAAATCGAACCTTATCTCTGAATTCATCTGCTTACC 1044
 Db 312 -----ArgGluMetSerIleAspAsnThrValPheLysSerThrLeuValAla 327
 QY 1045 CCTCTGGGAGAACTCATCTGTTACGTTGGGGGGGAGTTTCAGAGCTCGTCCATGAAA 1104
 Db 328 ProIleGlyGlu---HisMetLeuSerPheGlyValGluGlyLysHisGluSerLeuGlu 346
 QY 1105 GACGAGTGTCTCTCCACAGCAGGTGAAACT---TTCGGCGAGAAAGCTGGTCCGTA 1161
 Db 347 AspLysThrSerAsnLysIleSerSerArgTyrHisIleSerAsnThrGlnTrpAlaGly 366
 QY 1162 TTTGCTGAGGATCAGTGGCATCTCACGGATGCACTTGGCTGCTGCTGCGGAGCCGCTAT 1221
 Db 367 PheIleGluAspGluTrpAlaLeuAlaGluGlnPheArgLeuThrPheGlyArgLeu 386
 QY 1222 GAACATCATGAGCAATTCGGGGGACATCTCAGTCCCGTGCATATCTGCTGCGGATG 1281
 Db 387 AspHisAspLysAsnTyrGlySerHisPheSerProArgValTyrGlyValTrpAsnLeu 406
 QY 1282 GCAGATGCTTGCAGCTGAAAGCGGTGACACCGGATATAAGCCACCCAGATGGG 1341
 Db 407 AspProLeuTrpThrValLysGlyValSerThrGlyPheArgAlaProGlnLeuArg 426
 QY 1342 CAGCTACATAAAGGATTAGTGTGCTCGGCGAGGAAAAACAAATCTACTTGGTAAC 1401
 Db 427 GluValThrProAspTrpGlyGlnValSerGlyGlyGly-----AsnIleTyrGlyAsn 444

FT SIGNAL 1 25
FT CHAIN 26 663 COLICIN I RECEPTOR.
FT SITE 31 38 TONB BOX.
FT SITE 646 663 TONB C-TERMINAL BOX.
FT CONFLICT 97 97 S -> D (IN REF. 6).
FT CONFLICT 528 528 I -> N (IN REF. 1).
SQ SEQUENCE 663 AA; 7395 MW; 2C68A45D4B5EE414 CRC64;

Alignment Scores:
Pred. No.: 2,87e-52 Length: 663
Score: 832.50 Matches: 234
Percent Similarity: 47.45% Conservative: 120
Best Local Similarity: 31.37% Mismatches: 237
Query Match: 22.31% Indels: 155
DB: 1 Gaps: 25

US-10-625-972-4 (1-2091) x CIRA_ECOLI (1-663)

QY 40 CTCGGATTTCAGCCAGCAGCATAGCTGTGCA----- 72
Db ValGlyLeuCysLeuSerAlaIleSerCysAlaTrpProValLeuAlaValAspAsp 29
QY 73 GAGGATGTGATGTTCTCCGATCCGGCTATCAGAAAGCTGACTAACGCGCGCC 132
Db GlyGluThrMetValThrAlaSerSerValGluGlnAsnLeuLysAspAlaPro 49
QY 133 AGTGTCTCTGATTAGCCAGGAGGAATTCAGTCCAGCCAGTACCAGCATCTGGCGGAG 192
Db SerIleSerValIleThrGlnGluAspLeuGlnArgLysProValGlnAsnLeuLysAsp 69
QY 193 GCTCTGAGTACAGTAGAGGTGGATGTT---GAAAGTGGTACGGGTAAACCGGAGG 249
Db ValLeuLysGluValProGlyValGlnLeuThrAsnGluGlyAspAsnArgLysGly 88
QY 250 CTGGAATACAGATCCGAGGAATCCAGCCAGTACAGCTGATATGATGATGTT 309
Db ValSerIleArgGlyLeuAspSerSerTyrThrLeuIleLeuValAspGlyLys 106
QY 310 CGTCAGCGCGGAGCAGTACGTGACTCCCAACGGTTTTCTGCCATGATACCGGGTTC 369
Db ArgValAsnSerArgAsnAlaValPheArgHisAsnAspPheAspLeuAsn 123
QY 370 ATGCCCTCTGCGCCCATGAGCGTATTAGGTTATCAGGGGGCCGATGCCACACTG 429
Db TrpIleProValAspSerIleGluArgIleGluValValArgGlyProMetSerSerLeu 143
QY 430 TATGGCTCTGATCGATGGCGGTGGTGAATATCATTACCAGAAAGAAATCCAGACAA 489
Db TyrGlySerAspAlaLeuGlyGlyValValAsnIleThrLysLysIleGlyGlnLys 163
QY 490 TGGCTCTCTCCGTCATCAGCGGTGAATCTCAGGAAAGCAACAAATGGGGTAAACAGC 549
Db TrpSerGlyThrValThrValAspThrThrIleGlnGluHisArgAspArgGlyAspThr 183
QY 550 AGCCAGTTAATTTCTGGAGATGTGCCCTTGTGGATGATCTGTACGCTGCGAGTA 609
Db TyrAsnGlyGlnPhePheThrSerGlyProLeuIleAspGlyValLeuGlyMetLysAla 203
QY 610 CGCGGTAGCACACACACAGCT-----CAGGGTTCATCGGTACATCATCTG 654
Db TyrGlySerLeuAlaLysArgGluLysAspAspProGlnAsnSerThrThrAspThr 223
QY 655 AGGATACAGCAGCGCGGTATTCCTTATCCACGAGTCAAGAAATTATTAATCTTGGT 714
Db GlyGluThr----- 226
QY 715 GCACGCTTACCTGGAAGCGTCGGAGCAG-----GATGCTCTGGTTTGATATG 765
Db ProArgIleGluGlyPheSerSerArgAspGlyAsnValGluPheAlaTrp 243
QY 766 GATACACCGCGGAGCTTATGATAACCGGGATGGCAACTGGGAGTCTGACGGGGGGA 825
Db ---ThrProAsnGlnAsnHisAspPheThrAlaGlyTyr-----Gly 256

QY 826 TATGACCGG-----ACCCTGCGCTATGAGCGAAGCAAA 858
Db PheAspArgGlnAspAspAspSerLeuAspLysAsnAsgLeuGluArgGlnAsn 276
QY 859 ATTTCAGCTGGCTATGATCATCTTTCACCTTCGGAACATGGAATCGTATCTGAATCG 918
Db TyrSerValSerHisAsnGlyArgTrpAspTyrGlyThrSerGluLeuLysTyrTyrGly 296
QY 919 AACGAGACAGAAATAAAGGTCGTGAGCTTACGCGAGTGTACTGAAGCGCAGCAATGG 978
Db GluLysValGluAsnLysAsn----- 303
QY 979 GGGCTTCCGCTCAGCCCGCGGAGCTT---AAGGAATCGAACCTTATCTCGAATTCATTA 1035
Db ---ProGlyAsnSerSerProIleThrSerGluSerAsnThrValAspGlyLysTyr 321
QY 1036 CTGCTTACCCCTCTGGGAGATCTCATCTGGTACGGTGGGGGCGAGTTTCAGAGCTCG 1095
Db ThrLeuProLeuThrAlaIleAsnGlnPheLeuThrValGlyGlyGluTrpArgHisAsp 341
QY 1096 TCCATGAAGACGGAGTTGCTTCCGACGACAGTGAAACTTTCGCGCAGAAAGC--- 1152
Db LysLeuSerAspAlaValAsnLeu-----ThrGlyGlyThrSerSerLysThrSerAla 359
QY 1153 -----TGGTCGGTATTGCTGAGGATGAGTGGCATCTCAGGATGACACTTGGCTGACT 1206
Db SerGlnTyrAlaLeuPheValGluAspGluTrpArgIlePheGluProLeuAlaLeuThr 379
QY 1207 CGCGGACGCGCTATGAACATCATGAGCAATTTCGGGGACACTTTCAGTCCGCGTCATAT 1266
Db ThrGlyValArgMetAspAspHisGluThrTyrGlyGluHisTrpSerProArgAlaTyr 399
QY 1267 CTGCTCTGGATGTCGAGATGCTGACGCTGAAAGCGGTGTGACCCAGGATATAAG 1326
Db LeuValTyrAsnAlaThrAspThrValThrValLysGlyGlyTrpAlaThrAlaPheLys 419
QY 1327 GCACCCAGATGGGCGAGCTACATAAAGGATGAGTGTGTGTCGCGGAGGAGAAACA 1386
Db AlaProSerLeuLeuGlnLeuSerProAspTrpThrSerAsnSerCysArgGlyAlaCys 439
QY 1387 AATCTACTTGGTAAACCCGACCTGAAGCCGGAAGAGAGCGTTCAGTTATAGGCTGGGTG 1446
Db LysIleValGlySerProAspLeuLysProGluThrSerGluSerTrpGluLeuGlyLeu 459
QY 1447 TATTACGATAACCC-----GCCGCTCTGAATGCCAATGTCACAGGTTTATG 1494
Db TyrTyrMetGlyGluGluGlyTrpLeuGluGlyValGluSerSerValThrValPheArg 479
QY 1495 ACTGACTTCTCCACACAGATTGCTCTTATTCATAAATGATACACCAAT----- 1545
Db AsnAspValLysAspArgIle---SerIleSerArgThrSerAspValAsnAlaAlaPro 498
QY 1545 ----- 1545
Db GlyTyrGlnAsnPheValGlyPheGluThrGlyAlaAsnGlyArgArgIleProValPhe 518
QY 1546 AGCTATCTAAACAGCGGAAAGCGCGGTTCGACGGTGTGGAATTCGCGCACATTCGCCG 1605
Db SerTyrTyrAsnValAsnLysAlaArgIleGlnGlyValGluThrGluLeuLysIlePro 538
QY 1606 CTGTGGTTCAGAGGATGTCAGCTGTCTGCTGAATTAACCTGGACCCGAGTGAACACGT 1665
Db ---PheAsnAspGluTrpLysLeuSerIleAsnTyrThrTyr-----AsnAspGlyArg 555
QY 1666 GAT-----GCTGATAACAAAGTGGCGCTGAGTTATACCTGAAACAGCTGGTG 1716
Db AspValSerAsnGlyGluAsnLys-----ProLeuSerAspLeuProPheHisThrAla 573
QY 1717 AATCGGAACTGAATGGCAG-----ATCACGAAAGAGTGGCATCATGGCTGGGTGCC 1770
Db AsnGlyThrLeuAspTrpLysProLeuAlaLeuGluAspTrpSerPheTyrValSerGly 593

1771 CGTTATCCGGGAAACACACAGTTTCCACCAGAAATTATTCGTCACCTGAGCGCTGTACAG 1830
594 HistyThrGly-----Gln 598
1831 AAGAAAGTGTATGATGAGAAAGGAGAAVACCTGAAAGCTGGACGGTGGTGGATGCGAGGT 1890
599 LysArgAlaAspSerAlaThrAlaLysThrProGlyGlyTyrThrIleTpsAsnThrGly 618
1891 CTGTCGTGGAAGATGACGGATGCCCTGACCTGATGCTGCGGTGGAATAACCTCTCAAC 1950
619 AlaAlaTpsGlnValThrLysAspValLysLeuArgAlaGlyValLysLeuAsnLeuGlyAsp 638
1951 AAGGATTCACAGTACGCTGACGCTGACAGTGCCTGAGTACAGTGCCTGAGTACGCTGATGCGCGGTGAT 2010
639 LysAspLeuSer-----ArgAspAsp 645
2011 TACTCCAGACGGGATCATCAACACAGATATGTGATCTGACCTGACGCGGAATATCTGGATG 2070
646 Tyr-----SerTyrAsnGluAspGlyArgTyrPheMet 657
2071 TCGCTGAATCATCTTC 2088
658 AlaValAspTyrArgPhe 663
RESULT 3
FEPA ECOLI
ID FEPA ECOLI STANDARD; PRT; 746 AA.
AC P05825; P75722; P76821; P77093;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ferrienterobactin receptor precursor (Enterobactin outer-membrane
DE receptor).
GN FEPA OR FEP OR FEUB OR B0584.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86278160; PubMed=3015941;
RA Lundrigan M.D., Kadner R.J.,
RT "Nucleotide sequence of the gene for the ferrienterochelin receptor
RT FePa in Escherichia coli. Homology among outer membrane receptors
RT that interact with TonB.",
RL J. Biol. Chem. 261:10797-10801(1986).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.,
RT "The complete genome sequence of Escherichia coli K-12";
RL Science 277:1234-1238(1997).
RN [3]
RP SEQUENCE FROM N.A.
RA Chung E., Allen E., Araujo R., Aparicio A., Davis K., Duncan M.,
RA Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O., Lew H.,
RA Lin D., Namath A., Oefner P., Roberts D., Schramm S., Davis R.W.,
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.,
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome

corresponding to the 12.7-28.0 min region on the linkage map.";
DNA Res. 3:137-155(1996).
[5]
SEQUENCE OF 1-77 FROM N.A.
RX MEDLINE=89066678; PubMed=2974033;
RA Pettis G.S., Brickman T.J., McIntosh M.A.;
RT "Transcriptional mapping and nucleotide sequence of the Escherichia
RT coli fepA-fes enterobactin region. Identification of a unique
RT iron-regulated bidirectional promoter";
RL J. Biol. Chem. 263:18857-18863(1988).
RN [6]
RP MOLECULAR ANALYSIS.
RX MEDLINE=90354449; PubMed=2201687;
RA Armstrong S.K., Francis C.L., McIntosh M.A.;
RT "Molecular analysis of the Escherichia coli ferric enterobactin
RT receptor fepA";
RL J. Biol. Chem. 265:14536-14543(1990).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).
RX MEDLINE=99101384; PubMed=9886293;
RA Buchanan S.K., Smith B.S., Venkatramani L., Xia D., Esser L.,
RA Palnitkar M., Chakraborty R., van der Helm D., Deisenhofer J.;
RT "Crystal structure of the outer membrane active transporter FepA from
RT Escherichia coli";
RL Nat. Struct. Biol. 6:56-63(1999).
CC -!- FUNCTION: THIS PROTEIN IS INVOLVED IN THE INITIAL STEP OF IRON
CC UPTAKE BY BINDING FERRIENTEROBACTIN (FE-ENT), AN IRON CHELATIN
CC SIDEROPHORE THAT ALLOWS E.COLI TO EXTRACT IRON FROM THE
CC ENVIRONMENT. FEPA ALSO ACTS AS A RECEPTOR FOR COLICINS B AND D.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC -!- SIMILARITY: Belongs to the tonB-dependent receptor family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M13748; AAA65994.1; -
CC EMBL; AE000163; AAC73685.1; -
CC EMBL; U82598; AAB40783.1; ALT_INIT.
CC EMBL; D90700; BAA35225.1; -
CC EMBL; J04216; AAA23756.1; -
CC PIR; F64791; QRECFG.
CC PDB; 1FEP; 13-JAN-99.
CC SWISS-2DPAGE; P05825; COLI.
CC ECO2DBASE; D079.0; 6TH EDITION.
CC EcoGene; EG10293; fepA.
CC InterPro; IPR000531; TonB_boxC.
CC Pfam; PF00593; TonB_dep_Rc; 1.
CC PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
CC PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
CC Transmembrane; Outer membrane; Iron transport; Transport; TonB box;
CC Signal; Receptor; 3D-structure; Complete proteome.
FT SIGNAL 1 22
FT CHAIN 1 22 FERRIENTEROBACTIN RECEPTOR.
FT SITE 23 746 TONB BOX.
FT SITE 34 41 TONB C-TERMINAL BOX.
FT SITE 729 746 A -> R (IN REF. 1).
FT CONFLICT 152 152 MISSING (IN REF. 1).
FT CONFLICT 403 403
FT HELIX 40 44
FT TURN 45 46
FT TURN 48 49
FT TURN 50 54
FT STRAND 55 60
FT HELIX 67 71
FT TURN 68 73
FT TURN 72 76
FT TURN 75 81
FT STRAND 77 84
FT TURN 85 85


```

FT TURN 87 90
FT TURN 92 96
FT TURN 97 98
FT STRAND 101 103
FT TURN 104 108
FT TURN 109 110
FT STRAND 111 112
FT TURN 115 118
FT TURN 123 124
FT TURN 133 134
FT STRAND 138 140
FT STRAND 141 148
FT TURN 149 149
FT STRAND 150 153
FT TURN 154 155
FT TURN 157 158
FT STRAND 162 168
FT STRAND 176 186
FT TURN 189 190
FT STRAND 194 204
FT STRAND 209 220
FT TURN 225 231
FT STRAND 235 236
FT TURN 237 239
FT TURN 240 241
FT STRAND 249 264
FT TURN 265 266
FT STRAND 267 281
FT TURN 294 298
FT TURN 299 300
FT STRAND 304 318
FT TURN 320 321
FT STRAND 324 338
FT STRAND 360 378
FT STRAND 382 397
FT STRAND 426 441
FT STRAND 446 457
FT TURN 458 460
FT STRAND 461 473
FT TURN 476 477
FT STRAND 478 489
FT TURN 493 496
FT TURN 498 499
FT STRAND 501 503
FT STRAND 517 519
FT TURN 522 523
FT STRAND 527 540
FT TURN 541 542
FT STRAND 543 560
FT STRAND 567 567
FT TURN 569 570
FT STRAND 573 597
FT TURN 598 599
FT STRAND 600 614
FT TURN 615 617
FT STRAND 627 638
FT TURN 639 640
FT STRAND 641 650
FT STRAND 653 653
FT STRAND 658 668
FT TURN 660 661
FT STRAND 664 664
FT TURN 667 670
FT STRAND 673 673
FT STRAND 676 686
FT STRAND 691 698
FT TURN 700 701
FT STRAND 705 705
FT STRAND 714 715
FT TURN 717 718
FT STRAND 722 723
FT TURN 727 728
FT STRAND 733 733

FT STRAND 737 746
SQ SEQUENCE 746 AA; 82107 MW; 09348AAB1C29710A CRC64;

Alignment Scores:
Pred. No.: 2,73e-37 Length: 746
Score: 625.00 Matches: 209
Percent Similarity: 45.25% Conservative: 134
Best Local Similarity: 27.57% Mismatches: 285
Query Match: 16.75% Indels: 130
DB: 1 Gaps: 29

US-10-625-972-4 (1-2091) x FEPA_ECOLI (1-746)
QY 73 GAGGATGTGATGATGCTCGGCATCCGGCTATGAGAAAAGCTGACTAAGCAGCGGCC 132
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 33 AspAspThrIleValValThrAlaAala-----GluGlnAsnLeu-----GlnAlaPro 48
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 133 AGTGTCTTCTGTGATTAGCCAGGAGGAATTGCGAGCCAGCCAGTAC---CAGCATCTGGCG 189
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 49 GlyValSerThrIleThrAlaAspGluIleArgLysAsnProValAlaAaArgAspValSer 68
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 190 GAGGCTCTGAGATCAGTAGAGGGTGTGGATGTT-----GAAAGTGTGACGGGTAAACC 243
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 69 LysIleIleArgThrMetProGlyValAsnLeuThrGlyAsnSerThrSerGlyGlnArg 88
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 244 GGA---GGGCTGGAATCAGCATCCGAGGAATGCCAGCCAGTTCACGCTGACTACTGATT 300
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 89 GlyAsnAsnArgGlnIleAspIleArgGlyMetGlyProGluAsnThrLeuIleLeulle 108
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 301 GATCGT-----GTTCTGTCAGGGCGGGAAGCAGTACGTTGACT 336
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 109 AspGlyLysProValSerSerArgAsnSerValArgGlnGlyTrpArgGlyGluArgAsp 128
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 337 CCCAACGGTTTTCTCGCATGAATACCGGGTTCATGCCCTCTGGCCGCCATTGACCGT 396
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 129 ThrArgGly-----AspThrSerTrpValProPro---GluMetIleGluArg-143
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 397 ATTGAGGTATTACGGGGCGCGATGTCACACACTGTATGGCTCTGATGCGATGGCGGTGTG 456
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 144 IleGluValLeuArgGlyProAlaAlaAalaArgTyrGlyAsnGlyAlaAalaGlyGlyVal-163
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 457 GTGAATATCATTACCAAGAAATGACACAATGGCTCTCTCCGTCATCATGAGGGCTG 516
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 164 ValAsnIleIleThrLysLysGlySerGlyGluTrpHisGlySerTrpAspAlaTyrPhe 183
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 517 AATCTGCAGGAAGCAACAATGGGTAAACAGCAGCCAGCTTAATTTCTCGCAGCAGTGT 576
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 184 AsnAlaProGluHisLysGluGluGlyAlaThrLysArgThrAsnPheSerLeuThrGly 203
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 577 CCCCTTGTGGATGATCTGTCCAGCTCGAGGTACGGGTAGCAGACACACAGCGTCAGGGT 636
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 204 ProLeu---GlyAspGluPheSerPheArgLeuTyrGlyAsnLeuAspLysThrGlnAla 222
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 637 TCATCG-----GTCACATCACTGAGCGATACAGCAGCAGCGCGT 675
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 223 AspAlaTrpAspIleAsnGlnGlyHisGlnSerAlaAaArgAlaGlyThrTyrAlaThrThr 242
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 676 ATTCTTATCCACGGAGTCACAG---AATTATAATCTTGGTGCACGCTCTTACTGGAAG 732
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 243 LeuProAlaGlyArgGluGlyValIleAsnLysAspIleAsnGlyValValArgTrpAsp 262
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 733 CGCTCGGAGCAGGATGCTCTGTTTGATATGATACACCCCGGAG-----780
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 263 PheAlaProLeuGlnSerLeuGluLeuAlaGlyTyrSerArgGlnGlyAsnLeuTyr 282
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 781 -----CGTTATGATAACCGGGATGGCAACTGGGAGTCTGACGGGGGATAT 828
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 283 AlaGlyAspThrGlnAsnThrAsnSerAspSerTyrThrArgSerLysTyrGly-----300
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 829 GACCGGACCTCGCC---TATGAGCGAACAATAATT-----TCAGCTGGCTATGAT 876
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 301 AspGluThrAsnArgLeuTyrArgGlnAsnTyrAlaLeuThrTrpAsnGlyGlyTrpAsp 320
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
```

877	CATACCTTTCCACTTTCGGAAACATGGAATAATCGTATCTGAACCTGGAAACGAGACAGAAAATAA	936
QY	:::	
Db	AsnGlyValThrThrSerAsnTrpValGlnTyr-----GluHisThrArgAsn---	336
QY	937 GGTCTGTGACCTTGTCAGCAGTAGTACTGAAGCCGCAACAATGGGGCTTGCCTGTCAGCGG	996
Db	337 -----SerArgIleProGluGlyLeuAlaGlyThr	347
QY	997 CGG-----GAGCTTAAGAATCGAACCTT	1020
Db	348 GluGlyLysPheAsnGluLysAlaThrGlnAspPheValAspIleAspLeuAspVal	367
QY	1021 ATCTCGAATTCACTACTGTTACCCTCTCGGA-----GAATCtCATCTGGTTCAGGTG	1074
Db	368 MetLeuHisSerGluValAsnLeuProIleAspPheLeuValAsnGlnThrLeuThrLeu	387
QY	1075 GGGGGCGAGTTTCAGAGCTCGTCCATCAA-----	1104
Db	388 GlyThr-GluTrpAsnGlnGlnArgMetLysAspLeuSerSerAsnThrGlnAlaLeuThr	407
QY	1105 -----GACGGAGTTGCTTCCCGACGACACAGGTGAACCTTTC	1140
Db	408 GlyThrAsnThrGlyGlyAlaIleAspGlyValSerThrThrAspArgSerProTyrSer	427
QY	1141 CGGCAGAAAAGCTGGTTCGGTATTGCTGAGCATGAGTGGCATCTCACGGATGCACTTGG	1200
Db	428 LysAlaGluIlePheSerIleuPheAlaGluAsnAsnMetCulLeuThrAspSerThrIle	447
QY	1201 CTGACTCGGGCAGCGCTANGAACATCATGAGCAANTCGGGGACACTTCAGTCGGCT	1260
Db	448 ValThrProGlyLeuArgPheAspHisHisSerIleValGlyAsnAsnTrpSerProAla	467
QY	1261 GCATATCTGCTCGGATGNGGCAGATGCCCTGACGCTGAAGCGGTGACCAACGGGA	1320
Db	468 LeuAsnIleSerGlnGlyLeuGlyAspPheThrLeuLysMetGlyIleAlaArgAla	487
QY	1321 TATAAGGCACCCAGAATGGGCGACGTACAT-----AAAGGGATT	1359
Db	488 TyrLysAlaProSerLeutyGlnThrAsnProAsnTyriLeuTyrSerLysGlyGln	507
QY	1360 AGTGGTGTGTCGGGCGAGGAAAAACAAATCTACTTGGTAACCCCGACCTGGAAGCCGGA	1419
Db	508 GlyCysTyrAlaSerAlaGlyGlyCysTyrLeuGlnGlyAsnAspLeuLysAlaGlu	527
QY	1420 GAGAGCTCATGTTATGAGCTGGGGTGTATTACGATAACCCCGCGTTCGAATGCCAAT	1479
Db	528 ThrSerileAsnLysGluIleGlyLeuGluPheLysArgAspGlyTrpLeu---AlaGly	546
QY	1480 GTACAGCTTTTATGACTGACTTCTCAAACAGATTCTCTCT-----	1521
Db	547 ValThrTrpPheArgAsnAspTyrArgAsnLysIleGluAlaGlyTyrValAlaValGly	566
QY	1522 --TATTCCATAAATGATACACCAATAGCTGTATAACACCGGAAAGGCCCGGTTCAC	1578
Db	567 GlnAsnAlaValGlyThrAspLeuTyrGlnTrpAspAsnValProLysAlaValGlu	586
QY	1579 GGTGTGAATTTGCCGACATTCGCG-----CTGTGGTCAGAGGATGTC	1623
Db	587 GlyLeuGluGlySerLeuAsnValProValserGlnThrValMetTrpThrAsnAnile	606
QY	1624 ACGCTGTCACTGAATTACACCTGGACCGCAAGTGAACAACTGATGTGTATGATCAACAAGT	1683
Db	607 ThrTyrMetLeu-----LysSerGluAsnLys-----ThrThrGly	618
QY	1684 CGCGCGGTGAGTTATACCCCTGAACACATGGTGAATCGGAAACCTGAATCGCAGATCACC	1743
Db	619 AspArgLeuSerIleIleProGluTyrThrLeuAsnSerThrLeuSerTrpGlnAlaArg	638
QY	1744 GAAGAGGTGGCATCATGCTGGTCCCGTTTATCGCGGAAAAACACACCGTTTCACCCAG	1803
Db	639 GluAspLeuSerMetGlnThrThrPheThrThrTrpTyrGlyLysGlnGlnProLysLysTyr	658
QY	1804 AATTATTCTGCTCACTGAGCGCTGTACAGAGAAGAGTGTATGATGATGAGAAGAGGAATACCTG	1863

```

Db      659 AsnTy1LyGlyClnProAlaValGlyProGluThrLysGlu-----lle 673
QY      1864 AAAGCCTGGACCGTGGTGGATCAGCGTCTGTCTGTGGTGAAGATGACGATCGCCTGACGCTG 1923
Db      674 SerProTy1SerIleValGlyLeuSerAlaThrTrpAspValThrLysAsnValSerLeu 693
QY      1924 AATGCTGCGGTGAATAAATCACTGCTCAACAGGATTACAGTGACGTGAGCGCTGTACAGTGCC 1983
Db      694 ThrGlyGlyValAspAsnLeuPheAspLys-----ArgLeuTrpArgAla 708
QY      1984 GGTAAGAGTACGCTG-----TATGCGCGGTGATTACTTCAGCGGATCATCAACA 2034
Db      709 GlyAsnAlaGlnThrThrGlyAspLeuAlaGlyAlaAsnTy1IleAlaGlyAlaGlyAla 728
QY      2035 ACAGGATATGTGTATCACTGAGCGAAATTAATCGATGTCGCTGAACTATCAGTTC 2088
Db      729 Ty1ThrTy1AsnGluProGlyArgThrTrpTy1MetSerValAsnThrHisPhe 746

RESULT 4
PFEA_PSEAF
ID_PFEA_PSEAE STANDARD; PRT; 746 AA.
AC Q05098;
AD 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ferric enterobactin receptor precursor.
GN PFEA OR PA2688.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RN SEQUENCE FROM N.A.
RP RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=30437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Huftagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
CC CC -1- FUNCTION: SPECIFIC RECEPTOR FOR THE SIDEROPHORE FERRIC
CC ENTEROBACTIN.
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- INDUCTION: BY IRON AND ENTEROBACTIN.
CC -1- SIMILARITY: Belongs to the tonB-dependent receptor family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. See http://www.isb-sib.ch/announcement/
CC or send an email to license@sb-sib.ch.
CC
CC EMBL; M98033; AAA25928.1; -.
DR EMBL; AE004697; AAG06076.1; -.
DR F01; A40636; A40636.
DR HSRP; P05825; 1PEP.
DR InterPro; IPR000437; Prok_lipoProt_S.
DR InterPro; IPR000531; TonB_boxC.

```

